Topological Properties of Inequality and Deprivation in an Educational System: Unveiling the Key-Drivers Through Complex Network Analysis

Harvey Sánchez-Restrepo[1], Jorge Louçã[1]

[1] Information Sciences, Technologies and Architecture Research Center (ISTA), ISCTE-IUL, Lisbon, 1649-026, Portugal
{Harvey_Restrepo, jorge.l}@iscte-iul.pt

Abstract. This research conceives an educational system as a complex network to incorporate a rich framework for analyzing topological and statistical properties of inequality and learning deprivation at different levels, as well as to simulate the structure, stability and fragility of the educational system. The model provides a natural way to represent educational phenomena, allowing to test public policies by computation before being implemented, bringing the opportunity of calibrating control parameters for assessing order parameters over time in multiple territorial scales.

This approach provides a set of unique advantages over classical analysis tools because it allows the use of large-scale assessments and other evidences for combining the richness of qualitative analysis with quantitative inferences for measuring inequality gaps. An additional advantage, as shown in our results using real data from a Latin American country, is to provide a solution to concerns about the limitations of case studies or isolated statistical approaches.

Keywords: Complex Network · Educational Deprivation · Inequality · Large-scale assessments · Policy informatics

1 Introduction

Since the Sustainable Development Goals were launched [1], policymakers have been increasing the use of analytical and statistical models for improving their knowledge about complex dynamics of the educational systems, especially for those with high ethnic-linguistic diversity [2], facing the challenge of developing coherent multilevel theories for making decisions and designing public policies [3]. The wide number of nonlinear relationships exhibited by multiple agents in different hierarchical levels interacting in education, demands the development of new instrumental and thought tools for modelling the variety of fluxes of energy — financial, human, physical and social resources — as well as a better understanding and measurement of the parameters related with the emergence of self-organized patterns and groups in different scales [4].

Network theory has enormously developed in this decade and is a leading scientific field for descripting and analyzing complex social phenomena [5]. Educational systems exhibit different properties in many scales coming from same dynamics, students stablish relationships with other students and teachers, which might provide some insights about social characteristics about preferences, choices and interest on learning [6]. Actually, many interventions show that interactions impact instructional quality and learning outcomes [7] and there are some evidences on how social network structure becomes an important intermediate variable in education and that cultural, social and economic variables are related with educational deprivation and learning outcomes [8].

2 Modelling framework

In this research we incorporate a rich framework for analyzing topological and statistical properties of educational deprivation in a Latin American country, as well as its relationship with social determinants as Socioeconomic Status (SES), Rurality of area where the school is located (RA), Type of school (TS), and self-identify student's Ethnicity (ET), for unveiling the key factors driving inequality gaps in learning outcomes. The model is developed through a network with different levels for analyzing properties and nodes exhibiting centrality and non-equilibrium parameters than might help to better understand the structure of the system and phenomena behind them.

Data sources. A multivariate dataset related with learning outcomes of every student who has completed the *k-12* education process, estimated by scoring based on a census-based large-scale assessment carried out in Ecuador for 39 219 students in 2017, through a standardized computer-based test with psychometric parameters estimated by Item Response Theory with 2P-Logistic model. Raw scores were re-scaled to a *Learning index* (LI), a monotonous transformation of ability's parameter θ^i, where higher levels of learning are more likely to have higher scores.

The model was developed in four phases, the first one was psychometrical analysis for estimating scores and identify deprived students (L_0) — those with a LI below of the cut point L_0 — and the intensity of deprivation $\lambda(LI_i)$, given by the distance from LI_i to L_0. The next three phases are directly based on the level-network (LN). In this way, the 1-LN is for disaggregating L_0-group by SES, each student is represented by a node and edges are directed to one of the SES-decile nodes $\{\theta^i \to L_j^i \to (SES_k^i)\}\ \forall\, i$, weighted by λ. In 2-LN we extend 1-LN for including RA, TS and ET to analyze their effects through the sequence $\{\theta^i \to L_j^i \to (SES_k^i) \to (RA_k^i, TS_k^i, ET_k^i)\}\ \forall\, i$. The 3-LN amplifies and strengthens the network through more than one hundred educational and non-educational factors associated with learning achievements [9], trough the sequence $\{\theta^i \to L_j^i \to (SES_k^i) \to (RA_k^i, TS_k^i, ET_k^i) \to AF_{m,n}^i\}\ \forall\, i$, a multi-dimensional system exhibiting educational deprivation at different levels. Network analysis was carried out by Gephi 0.9.2 and statistical estimations with Orange 3.3.8.

3 Empirical results

Estimates indicate that, from 39 219 students, 8 438 were deprived, an absolute prevalence of 21.5%. The L_0-group has a $LI = 6.32$ and intensity of deprivation $\lambda = 0.22$, i.e., in average, each L_0-student lacks 0.68 standard deviations (SD) to the minimum level of learning for not being deprived.

3.1 Socioeconomic status and learning deprivation

In all cases, inequality means asymmetries, in conditions of total equity — where socioeconomic factors would not produce differences — we might expect equal distribution of L_0-edges over the network, but 1-LN specification detects SES effects in nodes grouping L_0-students by deciles, driving the system out of equilibrium with a negative correlation between LI and SES (R= –0.58 (p<0.001)). A 2-LN model shown in Figure 1 integrates the different self-identified ethnic groups and disaggregate them by Rural-Urban areas and Public-Private schools, to identify the magnitude with which the lower deciles dominate the interactions through the edges.

As can be seen in Figure 1, independently of RA, TS and ET, nodes corresponding to D01, D02 and D03 — the poorest students — have stronger connections and dominate the network, being D01 the highest parameters in Prevalence rate (*DPR*=0.3860), Hub (*H*=0.4887), Weighted Degree (*WD*=158.9420) and PageRank (*PR*=0.4289). On the contrary, D10 is almost irrelevant for the network with parameters *DPR*=0.0800, *H*=0.1133, *WD*=27.222 and *PR*=0.0289.

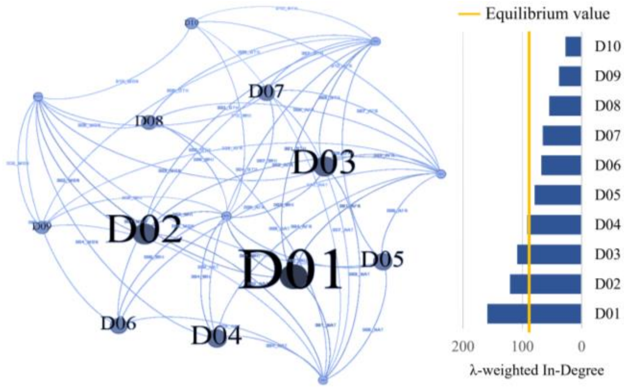

Fig. 1. Socioeconomic structure of learning deprivation (*left side*) and λ-weighted In-degree histogram (*right side*). Differential centrality of deciles shows a non-equilibrium L_0-system driven by ethnical groups dominated by poorest students.

3.2 Ethnicity and type of school financing

There is a negative correlation between SES and DPR (R= –0.81, p<0.001) and, as can be seen in scatter-plot of Figure 2, quintiles 1 and 2 have the highest lack of learning. The complementary Dendrogram was made by a hierarchical cluster analysis for

$\left(SES_k^i, RA_k^i, TS_k^i, ET_k^i\right)$ and, in all cases, families are basically made of SES and TS where emerges an intra-class ranking ordered by ethnicity, meaning that closeness is a SES' function while distances are based on racial proximities, a structural inequality for the whole system. In this sense, scatter-plot shows that White-students (red circle) have the lowest deprivation rate among the poorest (*DPR*=0.402), even lower than richest Afro-Ecuadorian students in blue circle (*DPR*=0.408). For the whole network, Page Rank order of nodes and λ-intensity is based on students' ethnicity as follows: Afro-Ecuadorian (A), Montuvios (M), Indigenous (I), Other groups (O) and White (W). Furthermore, Private sector is dominated by richest students (*SES*=0.91), while public schools serve to the poorest students (*SES*=0.33) getting a *DPR*=0.257, 2.12 times the rate of the private ones (*DPR*=0.121), showing that SES is a key factor for educational deprivation due to the influence of cultural capital in learning outcomes.

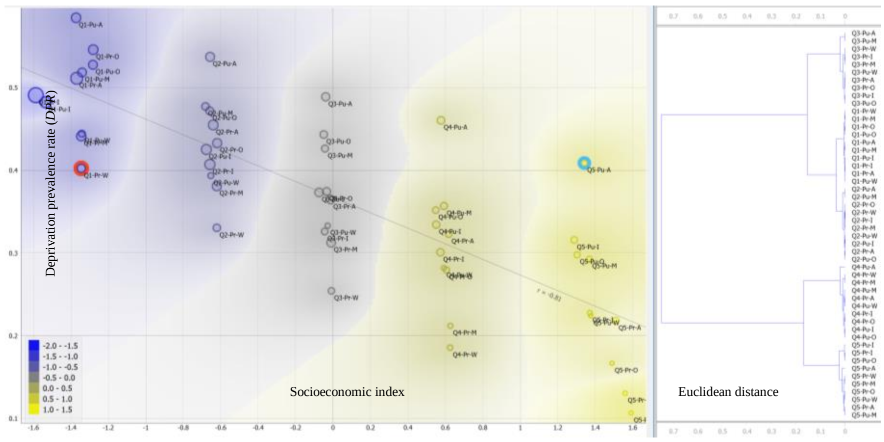

Fig. 2. Social determinants of learning deprivation. The scatter plot shows the strong negative correlation between SES and Deprivation prevalence rate (*solid line*). Highest deprivation rate in richest students (*blue circle)* is lower than the lowest rate of poorest students (*red circle*), pointing out a deep systemic racial discrimination.

3.3 Key-factors for public policies

As 1-Ln and 2-LN are networks strongly connected, prevalence rate of L_0-students might be associated with Eigen-Centrality through measuring factor's influence for identifying how well connected a node is and how many links have its connections. In this way, building 3-NL for splitting L_0-students in communities becomes in a very valuable tool for developing group-oriented strategies to avoid implementing same actions for completely different populations and needs. One of the most useful strategies for improving learning and closing gaps is micro-planning, i.e., implementing different policies at local level, however, to select the most relevant needs is a great deal, mainly because they use to interact and 'ceteris paribus hypothesis' seems to be too naive; through network simulation, this can be solved in a very easy way varying *AF*-parameters for recognizing and ranking the most relevant nodes (factors) to be attended by policymakers.

Figure 3 shows the 3-LN after a Modularity process (with parameter 0.073 at resolution of 0.254) for splitting richest from poorest students to find key factors for educational deprivation in both groups. As shown, richest students' community (D10) has just a few factors (20) and most of them have very low connectivity. On the contrary, poorest students (D01) have a lot of different sources provoking deprivation (57), reflected in the Degree of authority of the D01 (0.98743) versus D10 (0.158058).

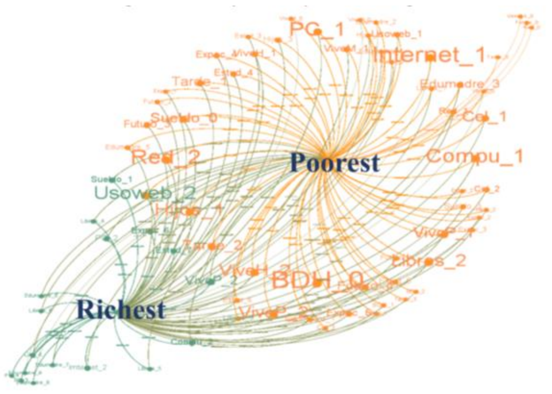

Fig. 3. Factor's network for richest and poorest students' communities. Centrality of nodes in each group points out the dissimilarities in factors provoking deprivation.

Thus, PageRank in 3-NL network helps to find and rank those factors dominating deprivation with the highest Degree of connectivity and centrality for each community ordering from the highest degree until the lowest for providing a very accurate and clear knowledge. In addition, it might be defined 'needs profiles' for focusing actions on those variables susceptible to be managed by policy, for example, a D01-profile might be: '*Members of a household who currently receives the Human Development Bond and needs to work for a wage. Their parents have a very low level of education, they do not have a desktop computer neither Internet connection, and also have no books or just a very few. In their school, teachers arrive late to class, are no committed with learning and have low expectations about student's future*'. As can be seen, this very detailed information is extremely helpful for developing based-evidence policies, to assign budget and have a successfully deployment.

4 Discussion

This approach provides a natural way to represent educational systems, bringing the opportunity of calibrating control-parameters for assessing order-parameters over time and in multiple territorial scales. Thus, the network analysis underpins current educational deprivation models [10] and provides a parametric way to estimate prevalence of inequality in learning outcomes for contexts with high levels of ethnic- diversity, a key aspect for understanding complexity in intercultural systems that also provides a solution to concerns about the limitations of case studies, the classical utility theory and isolated statistical analysis.

The findings offer evidence in the deep lack in equity that can help policymakers to identify those factors related with inequality in learning outcomes, as well as the magnitude of their relationship with deprivation. In this sense, running the modularity process for defining groups might help to identify those factors which are more relevant for one group and are not for others, avoiding statistical bias based on averages provoked by statistical multilevel modelling and bringing additional information about the order in which factors should be considered. Additionally, each LN can be stressed for calibrating boundaries and initial conditions, as well as to test policies at different levels, from-bottom-to-top and from-top-to-bottom and with real data, before being implemented.

These results confirm that network analysis is becoming fundamental for educational policy, specially linking microdata with other constructs and social macro-parameters. Finally, detecting how gaps in educational achievements are driven by students' context might highlight in a better way where policy can intervene properly, offering a series of unique advantages over classical analysis tools and allowing the intensive use of large-scale assessments and other datasets.

References

1. UN DESA. The Sustainable Development Goals Report 2018, UN, New York (2018)
2. Shee, N. K. Karen Education Department's multilingual education for language maintenance, Kasetsart Journal of Social Sciences. Elsevier (2018)
3. Johnson, Jeff. Hypernetworks in the science of complex systems. Imperial College Press, UK, London (2015)
4. Sánchez, Harvey. Equity: the focal point of educational quality. National Educational Evaluation Policy Gazette, Year 4. (10), pp. 42-44 (2018)
5. M. Kivela, A. et al. Multilayer networks, Journal of Complex Networks, vol. 2, no. 3, pp. 203–271 (2014)
6. Mizuno, K., Tanaka, M., Fukuda, S., Imai-Matsumura, K., & Watanabe, Y. Relationship between cognitive functions and prevalence of fatigue in elementary and junior high school students. Brain and Development, 33(6), 470-479 (2011)
7. Price, H. E., & Moolenaar, N. M. Principal-teacher relationships: foregrounding the international importance of principals' social relationships for school learning climates. Journal of Educational Administration, 53(1) (2015)
8. Najera Catalan H. E. Multiple Deprivation, Severity and Latent Sub-Groups: Advantages of Factor Mixture Modelling for Analysing Material Deprivation. Social indicators research, 131(2), 681–700 (2016)
9. OECD. Equity in Education: Breaking Down Barriers to Social Mobility, PISA, OECD Publishing, Paris (2018)
10. Gray, J., Kruse, S., & Tarter, C. J. Enabling school structures, collegial trust and academic emphasis: Antecedents of professional learning communities. Educational Management Administration & Leadership, 44(6), 875-891 (2016)

Acknowledgements

We would like to thank ISCTE's Programme in Complexity Sciences, the University of Lisbon and the Latin American Agency for Evaluation and Public Policy, for their support to the project as well as the authors of the articles mentioned in the references.