A Complex Network Approach to Structural Inequality of Educational Deprivation in a Latin American country

Harvey Sánchez-Restrepo [1[0000-0001-7311-8622]] and Jorge Louçã [2[0000-0003-4766-5627]]

[1] Faculty of Sciences, University of Lisbon, Cidade Universitária, 1649-004, Lisboa, Portugal
harvey@comunidad.unam.mx
[2] Information Sciences, Technologies and Architecture Research Center. ISCTE-IUL,1649-026, Lisbon, Portugal
jorge.l@iscte-iul.pt

Abstract. To guarantee the human right to education stablished by the fourth UNESCO's Sustainable Development Goal, a deep understanding of a big set of non-linear relationships at different scales is need it, as well as to know how they impact on learning outcomes. In doing so, current methods do not provide enough evidence about interactions and, for this reason, some researchers have proposed to model education as a complex system for considering all interactions at individual level, as well as using computer simulation and network analysis to provide a comprehensive look at the educational processes, as well as to predict the outcomes of different public policies.

The highlight of this paper is modeling the structure of the inequality of a national educational system as a complex network from learning outcomes and socio-economic, ethnicity, rurality and type of school funding, for providing a better understanding and measuring of the educational gaps. This new approach might help to integrate insights improving the theoretical framework, as well as to provide valuable information about non-trivial relationships between educational and non-educational variables in order to help policymakers to implement effective solutions for the educational challenge of ensuring inclusive and equitable education.

Keywords: Structural network, Large-scale assessments, Policy informatics.

1 Introduction

The 193 countries attached to Unesco promulgated the Sustainable Development Goals (SDG), the fourth goal (SDG-4) establishes that ''education is a human right' [1,2] and that the essential axes of quality in education must be learning and equity, since that all human beings have the right to learn, the State is obliged to guarantee the exercise of this right to all citizens equally [3,4]. However, Unesco points out that inequality in education has progressively increased and that the most impoverished bear the worst consequences [5,6]. In Latin America, the results of the last Large-scale Assessment of learning (LSA) show that most of the countries have high percentages of children with low-level learning after several years of schooling [7].

At global level, the lack of quality in education is of such magnitude that Unicef estimates that 250 million children, after having attended school, fail to develop the minimum learning in Literacy and Numeracy, both necessary to continue learning at the following educational levels [8]. For facing this challenge, many governments have promoted multiple reforms to improve the quality of their education systems, however, the modest improvements in learning have been accompanied by huge inequalities between different population groups, raising many questions to the policymakers [9,10].

In this context, to design better public policies, some countries have created national evaluation systems for measuring outcomes results and a pool of associated factors to learning (FAL) [4]. Most of those systems used standard tests based on psychometrical analysis and statistical models oriented to recognize the factors with a significant covariation, the main idea is estimating the average educational gains that a system might experience by varying one factor at a time through a specific policy [11,12]. Of course, the connection between constructs and observations has been a constant hypothesis of multiple experimental researches, however, one of the main questions to these actions is that, in daily practice, a fragmented vision of the system predominates, in which the interactions between the factors and the educational phenomena in their different scales are not considered [13,14].

Like most social systems, the educational ones exhibit non-linear relationships among their multiple agents with different levels of organization and time horizons that allow the emergence of self-organized phenomena, producing dynamic equilibria at different scales. However, dominant models for studying educational phenomena are based on reductionist tools that postulate that educational systems can be understood as the sum of their constituents where interactions among them are irrelevant. Therefore, the observed phenomenon is understood and modeled using one-to-one covariations for explaining lack or improvements in learning outcomes, dismissing the information coming from interactions between variables and their relevance in the structure of the entire system. Despite the historical success in some social research, in education this approach is, most of the times, just descriptive and lack of explanatory power [15].

In the last years, models based on Network science have been proposed as an alternative representation of systems, overall, because many systems can be described by complex interconnected networks as a result of self-organized processes [16]. One of the main advantages of modeling educational systems as a network, or multiple networks, is establishing connections between population characteristics of individuals as random phenomena with probabilities of occurrence given by data. This approach allows the identification of the factors that influence learning based on both topological and statistical parameters for unveiling some hierarchical structures related with inequality and educational deprivation.

For providing a robust model for better understanding inequality gaps, in this research we use concepts from network science as a tool for studying complexity and global and local properties of the structure of inequality in learning outcomes observed in a Latin American country. The analysis is based on statistical properties of the networks related with low-level-of-proficiency students and the Socio-economic Status

(SES) of the student's family, Rurality of the area where the school is located (RA), Type of school (TS), and Ethnicity (ET) for analyzing out-of-equilibrium states [19].

1.1 Dataset

For developing the model, a multivariate dataset integrates learning outcomes of every student who has completed the *k-12* education process, estimated by the ability's parameter θ^j through a LSA carried out in Ecuador in 2017, using a standardized computer-based test[1] and integrated with a robust dataset with more than 240 variables coming from surveys to student's families and teachers. For building the scores, ability and psychometric parameters were estimated by Item Response Theory as usual, through a 2P-Logistic model [19,20], following equation 1:

$$P(\theta_j) = \frac{e^{[\alpha_i(\theta_j-\beta_i)]}}{1+e^{[\alpha_i(\theta_j-\beta_i)]}} \text{ with } \theta_j, \alpha_i, \beta_j \in (-\infty, \infty) \tag{1}$$

After estimation process, raw scores were re-scaled to a standardized Learning index ($LI_j \in [4.0,10.0]$), a monotonous transformation of θ^j, where higher levels of learning are more likely to have higher scores [19,20]. The scores are on a continuous scale corresponding to four levels of achievement, according with the LSA and national standards, according with a technical description[2], all students are classified in Levels of Achievement (LA) stablished by a Bookmark process carried out by an expert pedagogical group on each subject [19]. The three psychometrical cut points s_1, s_2, s_3 correspond to four LA, where L_0 corresponds to those who have not reached a minimum level of learning, L_1 to the minimum acceptable, L_2 at the level of achievement raised by the system and L_3 corresponds to a performance higher than the standard.

1.2 Deprivation learning index

For estimating the Deprivation Learning Index (DLI), we use the family of scores $\{LI_j\}\ j = 1, \ldots, N,$ of those students with low level of achievement L_0-class where s_1 is the first cut point —the minimum score to be located at level L_1. They are students suffering learning deprivation according with the sociological proposal that states 'there is an irreducible nucleus of needs that are common to every human being', while relative deprivation, estimated by LI_j for each student, becomes from 'needs, thresholds and satisfactions are determined by each society' [21], both established by the Bookmark process.

For the L_0-class, absolute deprivation is given by H=$(n(L_0)/\sum n(LI_j)$, where $n(LI_j)$ represents the number of students below the first LA, the intensity $\lambda(LI_j)$ is given by the distance to reach the first level L_1, then DLI is given by δ_j=H$\cdot\lambda(LI_j)$, which

[1] Full dataset is available in http://www.evaluacion.gob.ec/evaluaciones/descarga-de-datos/, selecting the option *Ser Bachiller 2017-2018* and *microdato* for downloading the full data.

[2] Technical and pedagogical details about design can be found in http://www.evaluacion.gob.ec/evaluaciones/ser-bachiller/

represents a measure of the collective learning deficit, which considers the magnitude —the number of students with low performance— and intensity —how much below the minimum performance level are located [21].

As can be seen, histogram in Figure 1 shows that 22% does not meet the learning minimums at the end of the compulsory cycle.

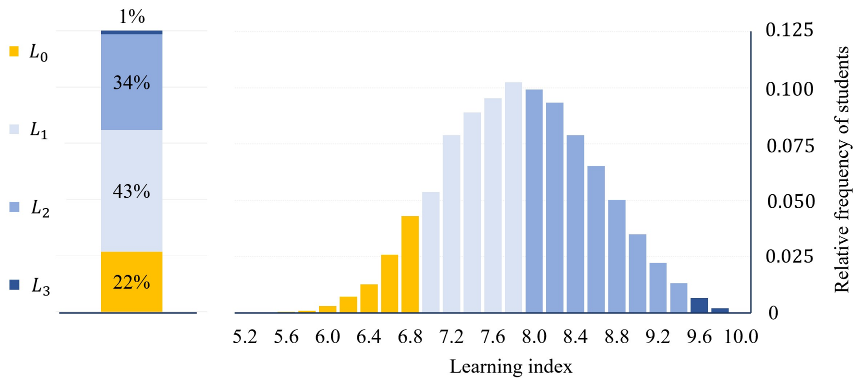

Fig. 1. Distribution of the students among Learning index (scores).

Scores distribution in Figure 1 allow the study of equity and it can be deepened by analyzing the levels of deprivation —absolute and relative— experienced by different population groups and their relationship with the socioeconomic status and ethnicity of students.

1.3 Model specification

In the last years, models based on graph theory have been proposed as a parallel representation of psychometric constructs such as intelligence, leadership or depression [22]. These models have in common that the covariance among the observable variables could be explained from the identification of patterns found among a set of interactions between these variables, measured through a set of informative items. The model for building the network is based on interactions between two nodes, representing the level of learning of each student directed to the set of each factor categories, where the edge is weighted by $\lambda(LI_i)$ [23].

For carrying on the analysis, the model runs in three phases: 1) analyzing the scores for assigning a LA to each student for identifying those located in level L_0, 2) estimating the SES for aggregated levels and subpopulation groups using the cut points for splitting in deciles, and 3) analyzing the associated factors to learning for creating the family of sequences $\{\theta^j \to L_k^j \to (SES_d^j)\}\ \forall\, j$ [24,25] for each student, where L_k^j is the LA, and SES_d^j corresponds to the SES decile of the j-th student.

To carry out a refinement of the variables that increase the deficit of basic skills in the population, the SES and the estimated deciles for the general population are preserved during all the stages, all other estimates are made again over the L_0 −group.

As each student is represented by a node, a set of edges, weighted by $\lambda(LI_j)$, are first directed to one of the SES-decile nodes $\{\theta^j \rightarrow L_k^j \rightarrow (SES_d^j)\}\,\forall\, j$, a process which allows to analyze aggregated inequality at school level, as well as In-degree distribution for SES nodes.

For extending the model and knowledge about social determinants, TS, RA and ET are included in the analysis one by one for analyzing their effects through the sequence $\{\theta^j \rightarrow L_k^j \rightarrow (SES_d^j) \rightarrow (TS_{C2}^j, RA_{C1}^j, ET_{C3}^j)\}\,\forall\, j$, where C denotes an index for each subcategory of the factors RA, TS and ET. Network analysis was carried out by Gephi 0.9.2 and statistical estimations and plots with R 3.5.0 and Orange 3.3.8.

2 Socioeconomic status and student's learning outcomes

To estimate the size of the gap at the macro level, the first network shown in Figure 2 integrates the Weighted In-degree distribution of directed edges from nodes indicating subpopulation groups to those representing SES deciles, given by $\{SES_d\}\; d \in \overline{1,10}$, where each edge represents one student in L_0-class. As inequality implies 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 for all deciles, but the distribution is not like that. Therefore, the study of equity can be deepened by analyzing the levels of absolute and relative deprivation experienced by different population groups and their relationship with the SES of the students.

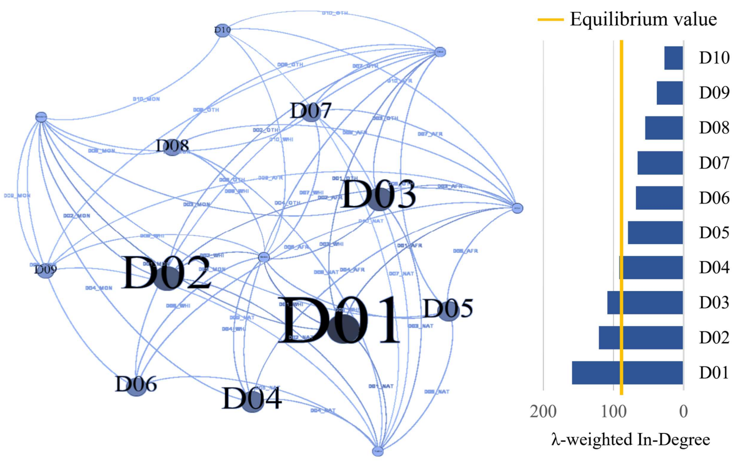

Fig. 2. Network for Weighted Out-degree and its edges' histogram shows socioeconomic status distribution of deprived students (L_0-group).

According with estimates, 21.5% of students are in L_0-class, a prevalence rate of 0.215 corresponding to a $LI = 6.32$ and shows an intensity of deprivation λ=0.225, i.e., in average, L_0-student lacks 0.68 standard deviations (SD) of the minimum learning.

To estimate the size of the gap at the macro level, Figure 2 also shows the percentages of students in each level of achievement for each SES decile. As can be observed, the proportion of students in each decile decreases monotonically as the SES of the group increases, being for the first decile (D01), 39% of students, and only 8% in D10. This difference of 31 percentage points is equivalent to the fact that for each rich family student who does not learn the minimum, there are 5 poor in the same situation. As will be shown later, this situation deepens in rural areas, where the ratio increases to one rich student for every 7 poor students.

3 The relationship between Type of school and SES

When studying schools as integrated units, the impact of SES becomes even more evident, in Figure 3 each school is represented by a circle whose size is proportional to the number of its students enrolled, the source of funding is distinguished by the color: green for private, blue for public. The average SES of students is located on the horizontal axis and the average score of LSA on the vertical axis. In the right side, the two whisker-box plots show dispersion for both indexes.

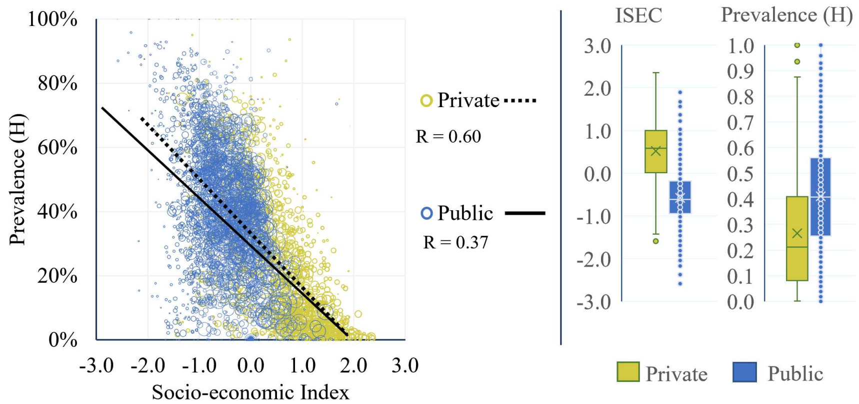

Fig. 3. Relationship between learning prevalence of deprivation and socioeconomic status at school level and box plots disaggregated by type of funding.

The negative correlation between SES and prevalence rate (R= —0.55, p<0.001) shows the separation of SES classes in groups of students who have different learning opportunities inside and outside schools, which helps to understand how inequality is gestated in a structural way in the country: schools with high SES predominate in the private sector and it is also there that the lowest levels of deprivation are presented, in this sector the correlation coefficient between the SES and H index is (R= —0.60, p<0.001). On the contrary, public schools that serve the poorest students have higher prevalence rates and a lower correlation (R = —0.39, p <0.001), which could indicate that, a weight of which the deprivation of learning is higher for the whole group, there is less inequality motivated by the socioeconomic origin of the student.

The socioeconomic gap between private (0.64) and public (—0.18) schools accumulates 0.82 standard deviations (SD), in addition, the prevalence rate in the public sector (*H*=0.219) is 1.7 times that of the private sector (*H*= 0.374). So, this confirms that public schools not only serve the poorest students in the country, but as they do so in the most depressed and most difficult areas, attendance is an extra challenge reflected in prevalence rates, while private schools concentrate on students in the top quintiles and the prevalence in most cases does not exceed 30%.

4 Inequality gaps and marginalized population groups

For having a more detailed and in-depth analysis, a selection of the two opposite SES population groups were selected as attractor nodes in the network —deciles D01 (Green) and D10 (Pink)—. In Figure 4, the network integrates the different ethnic groups of the country, disaggregated by rural and urban areas. Both, the nodes and the labels, are proportional to the *Weighted Out-Degree*, and $\lambda(LI_j)$ is weighting the edges.

The representation is based on the Eigen Centrality to measure the influence of a factor in terms of the number of edges with the population groups and that appear as other nodes within the network [27,28]. This measure is very valuable because, by counting how well connected a node is, and how many links have its connections through the network [29], the preponderance of the factors in the population groups becomes very clear in identifying the effect of the three educational deprivators: 1) rural areas, 2) types of funding of schools, and 3) ethnicity.

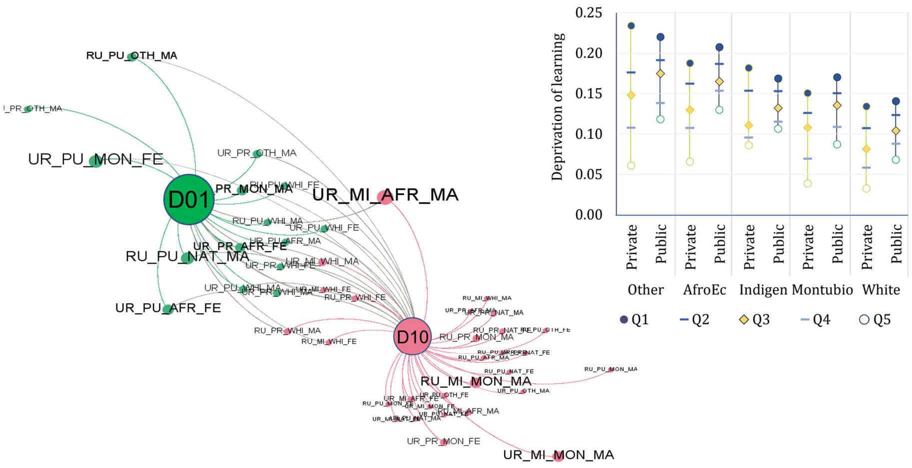

Fig. 4. Network of subpopulation groups splatted in rich and poor students with λ-weighted learning scores and its gaps synthesis.

For complementing information, the plot in the right side of Figure 4 shows that gaps between public and private sectors are quite pronounced, especially among the poorest students in both systems. In addition, although in the public sector there is less variability among quintiles, in all cases, Q1 and Q2 suffer the highest levels of deprivation. As can be seen, the lack of learning is found in the Afro-Ecuadorian population, Montubio's people, indigenous and other groups, before than people identified as White, the gaps are greater than 40% and show that a concentration in these historically vulnerable groups, mainly in the Montubios and Afro-Ecuadorians from rural areas, as well as indigenous people who attend school in urban areas.

Of special interest is the case when the modularity is represented, in this network, the parameter estimated was 0.073 and resolution -0.254, which produces two communities: the richest and the poorest. This result is particularly important for public policies because it might allow policymakers to work directly with families in a group-oriented strategy to avoid presenting same actions for completely different problems.

It is also remarkable that private schools serving indigenous and other minority groups coming from Q1, show higher deprivation rate than graduates of the public system in the same Q1 level. This result points out a tremendous social deception and suggests an urgent migration of those students to the public system in order to review the operation of these schools, given their low performance in a group with so many disadvantages. A racial dramatic case is also found: Afro-Ecuadorians with the lowest level of deprivation are those located in Q5, however, the poorest white students attending public schools show a level of deprivation equivalent.

This approach allows to measure the magnitude with which the lower deciles dominate in the interactions with the population groups through the edges [26]. Furthermore, the network has directed edges and its average weighted Out-Degree might be seen as a covariation-measure of the deciles with population of non-learning students, so, it is possible to compare the values thrown by the network with the value that could be expected on this parameter in conditions of equity where the factors would not produce differences.

5 Discussion

With this new kind of analysis, we have developed a model for finding answers to classical questions in educational research using free available data for a Latin American country, providing a direct method to recognize the structure of inequality, as well as the relationship between social determinants for educational deprivation and the conditional distribution of learning outcomes.

Given that equity is a major focus of government policies around the world and that it is promoted by international agencies with the aim of transforming educational systems, attending the wide diversity of students in each country and the whole region is a big challenge and in this paper we have presented an analysis that offers a lot of valuable evidence showing the deep lack in this dimension, highlighting that is a structural problem that goes beyond educational policy.

Using the network concept of modularity in defining groups with the same kind of challenges, might help to policymakers in selecting those factors that might be so much relevant for one group than to others, overall in some areas where the intra-class variability is very low using complementarian topological and statistical analysis.

Given the relationship between the DLI and SES, and that deprivation is almost eight times higher for poorest than for richest students, it is confirmed that the exercise of educational rights is a function of SES and that the gap is wider when considering the types of financing and that this phenomenon gets even worst for minority ethnic groups.

Extending the model for including a large set of factors should be the next step for improving the analysis and offer useful information about the educational system, as well as developing based-evidence public policies. This can be achieved introducing a dataset at micro level for building the network for meso and macro levels, a very challenging task when considering the number of students in a national education system and all the connections they may have with educational and non-educational variables. However, this point is crucial for policy because the social deception of accessing to a school without the guarantees of learning translates in levels of precariousness similar to those who never attend the school [30], creating circles of poverty that systematically impact to the impoverished people, increasing the gaps and structural inequality.

References

1. OECD (2018). Equity in Education: Breaking Down Barriers to Social Mobility, PISA, OECD Publishing, Paris
2. UN DESA (2018). The Sustainable Development Goals Report 2018, UN, New York
3. MCGRATH, Simon; NOLAN, Aoife. SDG 4 and the Child’s Right to Education. NORRAG NEWS 54, 2017, p. 122.
4. Sánchez-Restrepo, H. (2018). Equity: the focal point of educational quality. National Educational Evaluation Policy Gazette in Mexico, Year 4. (10), pp. 42-44.
5. United Nations Children's Fund (Unicef). The investment case for education and equity. DC, Washington: 2015
6. Keeley, B., & Little, C. (2017). The State of the World’s Children 2017: Children in a Digital World. UNICEF. 3 United Nations Plaza, New York, NY 10017.
7. Samman, E. (2018). SDG progress: fragility, crisis and leaving no one behind: report.
8. Gray, J., Kruse, S., & Tarter, C. J. (2016). Enabling school structures, collegial trust and academic emphasis: Antecedents of professional learning communities. Educational Management Administration & Leadership, 44(6), 875-891
9. Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación – LLECE (2018). Agenda 2018. http://www.unesco.org/new/en/santiago/press-room/newsletters/ newsletter-laboratory-for-assesSSent-of-the-quality-of-education-llece/.
10. Hopfenbeck, Therese N., et al. (2018). Lessons learned from PISA: A systematic review of peer-reviewed articles on the programme for international student assessment. Scandinavian Journal of Educational Research, vol. 62, no 3, p. 333-353.
11. Motl, T. C., Multon, K. D., & Zhao, F. (2018). Persistence at a tribal university: Factors associated with second year enrollment. Journal of Diversity in Higher Educ, 11(1), 51.
12. Moll, K., Göbel, S. M., Gooch, D., Landerl, K., & Snowling, M. J. (2016). Cognitive risk factors for specific learning disorder: processing speed, temporal processing, and working memory. Journal of learning disabilities, 49(3), 272-281.

13. Tracey, B., & Florian, K. (Eds.). (2016). Educational Research and Innovation Governing Education in a Complex World. OECD Publishing.
14. Fox, J. (2015). Applied regression analysis and generalized linear models. Sage Public.
15. Davis, B., & Sumara, D. (2014). Complexity and education: Inquiries into learning, teaching, and research. Routledge.
16. Barabási, Albert-László, and Márton Pásfai. Network Science. Cambridge: Cambridge University Press, 2016.
17. van der Maas, H., Kan, K., MarSSan, M., & Stevenson, C. (2017). Network models for cognitive development and intelligence. Journal of Intelligence, 5(2), 1–17. doi: 10.3390/jintelligence5020016
18. Johnson, J., Fortune, J., & Bromley, J. (2017). Systems, networks, and policy. In Non-Equilibrium Social Science and Policy (pp. 111-134). Springer, Cham.
19. Borsboom, D., & Molenaar, D. (2015). Psychometrics. In J. Wright (Ed.), International encyclopedia of the social & behavioral sciences. Vol.19 pp.418–422.Amsterdam: Elsevier
20. LI, Feiming, et al. (2009). Model selection methods for mixture dichotomous IRT models. Applied Psychological Measurement, vol. 33, no 5, p. 353-373.
21. Bracho, T (2017), Índice de déficit en competencias ¿Avanzamos hacia la garantía del derecho a la educación? en Reformas y Políticas Educativas; No. 4, septiembre-diciembre 2017, FCE, México.
22. Marsman, D. Borsboom, J. Kruis, S. Epskamp, R. van Bork, L. J. Waldorp, H. L. J. van der Maas & G. Maris (2018) An Introduction to Network Psychometrics: Relating Ising Network Models to Item Response Theory Models, Multivariate Behavioral Research, 53:1, 15-35.
23. Costantini, G., Epskamp, S., Borsboom, D., Perugini, M., Mottus, R., Waldorp, L.J., & Cramer, A.O.J. (2015). State of the Art personality research: a tutorial on network analysis of personality data in R. Journal of Research in Personality, 54, 13–29.
24. Kossakowski, J.J., Epskamp, S., Kieffer, J.M., van Borkulo, C.D., Rhemtulla, M., & Borsboom, D. (2016). The application of a network approach to health-related quality of life (HRQoL): introducing a new method for assessing HRQoL in healthy adults and cancer patient. Quality of Life Research, 25, 781–92.
25. Forbush, K., Siew, C., & Vitevitch, M. (2016). Application of network analysis to identify interactive systems of eating disorder psychopathology. Psychological Medicine, 46(12), 2667–2677.
26. Gibson, L., & Koenig, A. (2012). Neighboring groups and habitat edges modulate range use in Phayre's leaf monkeys (Trachypithecus phayrei crepusculus). Behavioral Ecology and Sociobiology, 66(4), 633-643.
27. van Borkulo, C.D., Borsboom, D., Epskamp, S., Blanken, T.F., Boschloo, L., Schoevers, R.A., & Waldorp, L.J. (2014). A new method for constructing networks from binary data. Scientific Reports, 4(5918), 1–10.
28. Forbush, K., Siew, C., & Vitevitch, M. (2016). Application of network analysis to identify interactive systems of eating disorder psychopathology. Psych Med, 46(12), 2667–2677.
29. van der Maas, 2017)
30. Menefee, T., & Bray, T. M. (2015). Education in the Commonwealth: Quality education for equitable development. Commonwealth Secretariat.