Idiographic Learning Analytics:
A single student (N=1) approach using psychological networks

Mohammed Saqr [1,2], Sonsoles López-Pernas [3]

[1] *KTH Royal Institute of Technology, Stockholm, Sweden*
[2] *University of Eastern Finland, Joensuu, Finland*
[3] *Universidad Politécnica de Madrid, Madrid, Spain*

Abstract

Recent findings in the field of learning analytics have brought to our attention that conclusions drawn from cross-sectional group-level data may not capture the dynamic processes that unfold within each individual learner. In this light, idiographic methods have started to gain grounds in many fields as a possible solution to examine students' behavior at the individual level by using several data points from each learner to create person-specific insights. In this study, we introduce such novel methods to the learning analytics field by exploring the possible potentials that one can gain from zooming in on the fine-grained dynamics of a single student. Specifically, we make use of Gaussian Graphical Models —an emerging trend in network science— to analyze a single student's dispositions and devise insights specific to him/her. The results of our study revealed that the student under examination may be in need to learn better self-regulation techniques regarding reflection and planning.

Keywords [1]

Graphical Gaussian Models, Idiographic Learning Analytics, Network Science, Psychological Networks

1. Introduction

The growing field of learning analytics (LA) has drawn the attention of academics, researchers, and administrators who aspire to understand and optimize teaching and learning [1]. Over ten years of findings have brought immense insights to our attention. One of the most important lessons that we have learned is that context matters: models obtained in one context are barely transferable to other contexts [2]. Researchers have failed to replicate the results of predictive models (e.g., for estimating student performance) across multiple learning settings due to the remarkable diversity in the data generated by students' learning activates, the obtained predictors, as well as the levels of statistical significance [3,4]. These inconsistencies have made the efforts towards offering adaptive learning or personalizing support an arduous endeavor. Researchers have called for using the high resolution data generated by students to generate *personalized* insights [5]. However, analyzing cross-sectional (i.e., group-level) data to generate *personalized* recommendations does not mean that each individual person will conform to the group average, and consequently, such insights generated by averaging over a group are hardly transmutable to every individual person [6]. Furthermore, cross-sectional group-level data fail to account for the dynamic processes (e.g., cognition and communication) that unfold within the individual. Obviously, a single cross-sectional timepoint is hardly useful to explain a dynamic phenomenon occurring over multiple time points [7].

On this basis, idiographic methods have started to gain grounds as a possible solution to examine behavior at the individual level in other fields. Idiographic methods use several data points from an individual to create person-specific insights. Being derived on the person level, such analyses account for the individual factors while being able to explain dynamic phenomena [8–10]. Winne et al. (2017)

Proceedings of the NetSciLA21 workshop, April 12, 2021
EMAIL: mmas3@kth.se (A. 1); sonsoles.lopez.pernas@upm.es (A. 2)
ORCID: 0000-0001-5881-3109 (A. 1); 0000-0002-9621-1392 (A. 2)

argued that high resolution data enable individual (i.e., idiographic) learner analytics, so that learners can gather own data and “interpret results to decide whether and how to adapt study tactics and learning strategies”. Dawson et al. (2019) examined a large sample of students and tried early interventions aiming at prevention of dropouts. Their findings pointed to no effect on the retention outcome. The authors concluded that more data about individual differences are needed to better understand the retention process as well as to design relevant *personalized* interventions. A recent massive scale study that has examined a large sample of students (around 250,000) have found small benefit of a group-based behavioral intervention despite the massive dataset. Authors concluded that the field needs efficient interventions tailored to the *individual* and course context. Thus, education researchers need to explore such individual-based approaches [7].

This study builds on the aforementioned insights and takes inspiration from the emerging fields of idiographic psychology and precision medicine, which have developed methods and standards for such methods of analysis [7,8,11]. In doing so, we explore the possible potentials that one can gain from zooming in on the fine-grained dynamics of a single student. We explore a person-specific data collection method as well as person-specific analysis and recommendations. Using data from a single student over 30 days, we analyze his/her dispositions and devise insights specific to him/her. Our approach is based on the emerging trend in network science, in particular, Gaussian Graphical Models (GGM) [10,12]. Our research question is as follows: What insights can idiographic learning analytics reveal about students’ self-regulation and learning dispositions?

2. Background

2.1. The cognitive process as a networked system

Representing elements of the cognitive and social processes as a network is an established research method. Such representation has afforded researchers a way to visualize the structure of these processes to measure the magnitude of association between their elements, and to devise statistical indices that allow a precise interpretation of the resultant graphs [13]. In education, research on networks spans three decades. Networks have been used to visualize the patterns of interactions in collaborative groups, to study the roles students play in the collaboration, to rank students’ activities, or to predict performance to mention a few examples [14–17]. While such methods have contributed enormously to our understanding of the learning process with their repertoire of powerful visualizations methods, there is a need for harnessing the power of other methods to extend our understanding different phenomena.

2.2. Gaussian Graphical Models

Recent advances in network sciences have led to the remarkable growth of probabilistic network models, often referred to as GGM [10]. GGM map the dynamic relationships between the elements of the cognitive or sociological phenomena we seek to understand as a complex system through the estimation of a network where the nodes are variables and the edges are the partial correlation coefficients between these variables [10,18–21]. Similar to multiple regression, partial correlations estimate the correlations after controlling for all other variables in the network, thus eliminating the possible effect of confounding variables [19]. This is particularly useful when there are multiple dependencies, i.e., consider an example when a researcher finds a positive correlation between coffee consumption and academic performance, such a correlation may simply be an unmeasured confounding factor (e.g., study time that leads to more coffee drinking). Thus, in GGM networks, two nodes are connected —if and only if— there is a covariance between these nodes that cannot be explained by any other variable in the network [10,12,18]. The resulting networks show only the significant relationships, the strength of such relationships, the sign (positive or negative), as well as the mediation pathways. Such rigorous network models offer “hypothesis generating structures, which may reflect potential causal effects to be further examined” [18]. As such, GGM offer several advantages that overcome the shortcomings of existing methods in terms of rigorous inferential statistics, ability to control for confounding factors, modelling the temporal evolution of the studied process. Moreover, there is a diverse and large community working on refining and improving GGM methods.

2.3. Graphical Vector Autoregression

An extension of GGM methods has allowed for the modeling of temporal processes, i.e., how a variable predicts another in the next time window. The abundance of intensive time-stamped data (time-series) has led to the existence of enough observations of individual subjects across short periods (e.g., experience sampling methods, observational data and physiological data), i.e., an individual can be studied as a unique case (N=1) [10,22]. Such time-series data are amenable to multivariate time-series analysis, commonly known as known as vector autoregression (VAR) [10]. VAR estimates a directed network (in contrast to undirected in GGM): the nodes are variables (e.g., motivation, behavior or attitude) and the link between them are temporal relationships (a variable predicts another in the next time window) [10]. This is commonly represented by drawing a directed arrow from the node that represents the variable (e.g., motivation) to the variable that it predicts in the next time window of measurement (e.g., engagement). An example is presented in Figure 1, which shows a temporal network generated from a fictional individual dataset about hourly eating and exercise habits. The graph illustrates that running predicts rest thereafter and that comfort predicts eating (weak prediction, see the thin line). The loop around comfort means that comfort at one hour predicts that the person will be at comfort the next hour; probably breaking the eating habits may entail keeping occupied with activities. As shown, a temporal network predicts if a variable (an element of the studied phenomena) predicts another in the next time window. Such type of network is used to explain within-subject covariation or potential causal pathways.

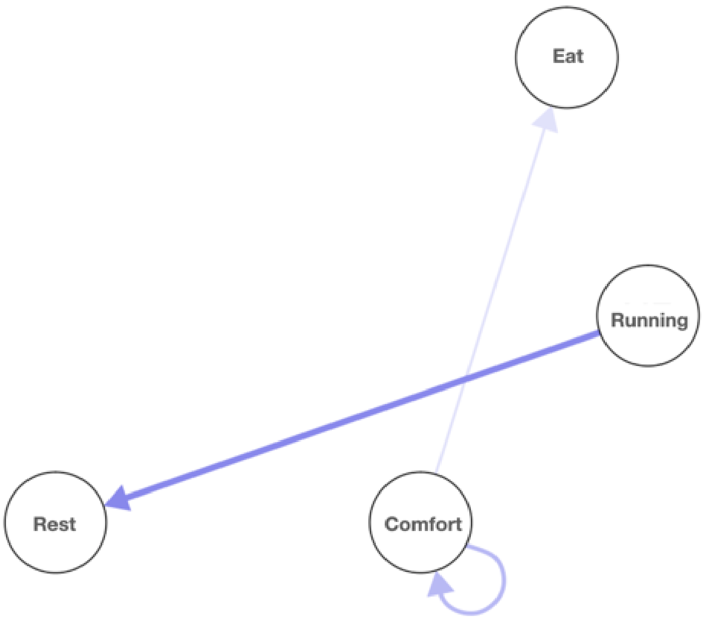

Figure 1: A fictional temporal network of four behaviors. The circles are variables. Blue lines are positive partial correlations. The thickness of the line is proportional to the magnitude of the correlation. The direction of the arrow points to the direction of the temporal correlation.

3. Methods

The study included a single student who signed an informed consent for an anonymous version of the responses to be used for research purposes. The student was attending a course over a duration of a month. The student had to respond to ten questions representing common dispositions and self-regulation (SRL) that are commonly employed in learning analytics [23–25]. The questions covered the following constructs: Expectancy value (Vlu), Motivation (Mtv), Stress as negative affect (Str), Hope and enthusiasm as positive affect (Hop), SRL Planning (Pln), SRL Engagement with task (Tsk), SRL

Reflection and evaluation (Rfc), External Regulation by assignments (Asg), Socializing (Soc), Challenging learning tasks (Chl).

The survey data was detrended using the method described in [26] to make the data close to stationary. Since our interest was to study the interplay between the student's different dispositions, we used the VAR model. VAR models have been established in the study of psychological phenomena, shedding light on the temporal progression, individual aspects and dynamics of psychological processes within individuals [10,26,27]. To understand the sequential temporal dependencies, we created a temporal network by estimating a Graphical VAR model [26]. The temporal network captures what will happen next as an effect of what is happening now (lag-1 or cross-lagged effects), e.g., if the person is motivated now, the person is going to work on the task on the next step. To account for multiple comparisons, the model was regularized using graphical least absolute shrinkage and selection operator (GLASSO). Using GLASSO algorithm for estimating GGM networks has been shown to retrieve the true structure of the network [26].

4. Results

The results of the temporal network showed interesting results about the involved student (Figure 2 and Table 1). After controlling for all other variables in the network, the positive affect (feeling hope) was the most predictive variable of engagement in a task in the next day, shown as a thick arrow between the *Hop* and *Tsk* nodes in Figure 2 indicating the strong association. Motivation was also strongly predictive of engagement with the task after controlling for all other variables, i.e., independent of feeling hopeful, socializing, etc. The challenging nature of the task was also predictive of engagement for the student, as well as stress, indicating that a bit of a challenge may help some students engage and work on the learning activities. The expected value and relevance of the task was also predictive of the student's engagement with the task, emphasizing the need for creating more relevant and authentic learning tasks.

Table 1
Values of the VAR partial correlations

	Tsk	Vlu	Mtv	Str	Hop	Pln	Rfc	Asg	Soc	Chl
Tsk	0.00	-0.02	0.00	0.00	0.00	0.00	0.00	0.01	-0.01	0.00
Vlu	0.16	0.00	0.03	0.00	0.00	-0.02	0.00	0.04	0.10	0.08
Mtv	0.27	0.03	0.00	0.03	0.00	-0.07	0.03	0.00	0.28	0.00
Str	0.17	-0.07	0.00	0.00	0.05	-0.01	0.00	-0.08	0.12	0.03
Hop	0.29	0.06	0.00	0.00	0.00	-0.05	0.05	-0.02	0.06	0.13
Pln	-0.06	-0.07	0.00	-0.06	0.00	0.00	0.00	0.00	-0.04	-0.09
Rfc	-0.22	0.01	0.00	0.00	0.00	0.00	0.05	0.01	0.00	0.00
Asg	-0.17	0.05	0.00	-0.09	-0.03	0.00	0.05	0.18	-0.17	0.01
Soc	-0.02	0.04	-0.01	-0.03	-0.02	0.02	-0.01	0.00	0.01	0.03
Chl	0.20	0.00	0.09	0.00	0.04	0.00	-0.09	0.00	0.00	-0.06

Working on the assignment was negatively predictive of engagement with learning tasks, as the student focused more on finishing the submissions. Such results also indicate that external regulation may be counterproductive for some students. Similarly, reflection was negatively predictive of engagement with the task the next day, which raises the question of the nature of reflection the student has. Planning was also weekly negatively associated with engagement with the task. These negative associations for assignment, reflection and planning are indicative of poor self-regulation practices by the student. In fact, the student had to repeat one of the assignments as it was not fulfilling the required guidelines and was incomplete. He also scored below the 50th percentile in the two most important course assignments. There is room for improvement here, by helping the student learn optimal self-regulation practices. There was a negative association between motivation and planning, while strong positive association with socialization. Stress and assignment negatively influenced each other: the more stress the student was under, the less he/she worked on the assignments, and the more work on assignments the less stressed was the student, as expected. The results are detailed in Table 1. Please note that, since partial correlations do control for other variables, their values are not to be interpreted in the same way, as they tend to be lower.

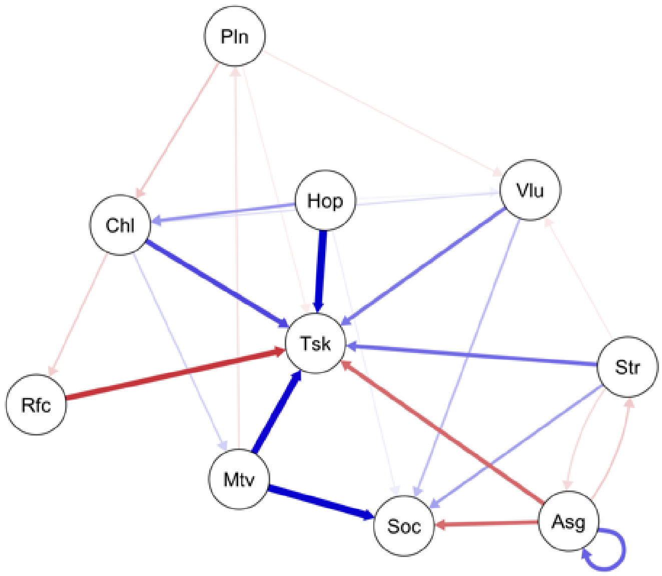

Figure 2: Temporal network for the student

5. Discussion and conclusions

In this study, we have used psychological network methods in the form of GGM and graphical VARs to study a single student disposition during a course. Such idiographic method offers several advantages over cross-sectional group level analysis. Being focused on a single student, the insights generated are more relevant and actionable, i.e., precisely personalized, paving the path for precision education. These methods also offer several advantages regarding controlling for confounders, deleting spurious correlations and regularization which requires high magnitude significant correlation, offering a good level of rigorousness [10,26,27]. The study has shown that the student under examination may be in need to learn better self-regulation techniques regarding reflection and planning based on his own responses. However, the value of such targeted intervention is yet to be investigated.

The implication of our study can be the applicability of the approach in several scenarios and contexts. Researchers who wish to apply personalized learning analytics can use such methods to design personalized intervention for their students. We believe there is an opportunity that may change the deserves attention and efforts from the research community to extend, improve and build on such methods. Our methods are not without limitations. The idea that the data have to be collected on a daily basis makes it sometimes difficult to collect data without some gaps, non-compliance, or missing values. The rate of data collection can be tricky: we have used a lag of a single day, but we do not know for sure if that lag was optimal. The timing of the data collection is another factor: whether data should be collected before or after the working day is still an open question. Similarly, how frequently data should be collected, what factors are to be included in the study, and how long we should collect the data are aspects in need of further investigation. The collection of data comes always with problems and risks of privacy and ethical concerns [28,29], in idiographic approach where much data is collected it can pose a risk which needs to be mitigated [30].

6. References

[1] G. Siemens, Learning Analytics: The Emergence of a Discipline, American Behavioral Scientist. 57 (2013) 1380–1400. https://doi.org/10.1177/0002764213498851.

[2] D. Gašević, S. Dawson, T. Rogers, D. Gasevic, Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success, The Internet and Higher Education. 28 (2016) 68–84. https://doi.org/10.1016/j.iheduc.2015.10.002.

[3] R. Conijn, C. Snijders, A. Kleingeld, U. Matzat, Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS, IEEE Transactions on Learning Technologies. 10 (2017) 17–29. https://doi.org/10.1109/TLT.2016.2616312.

[4] S. Dawson, S. Joksimovic, O. Poquet, G. Siemens, Increasing the Impact of Learning Analytics, in: Proceedings of the 9th International Conference on Learning Analytics & Knowledge, ACM, New York, NY, USA, 2019: pp. 446–455. https://doi.org/10.1145/3303772.3303784.

[5] P.H. Winne, J.C. Nesbit, F. Popowich, nStudy: A System for Researching Information Problem Solving, Technology, Knowledge and Learning. 22 (2017) 369–376. https://doi.org/10.1007/s10758-017-9327-y.

[6] A.J. Fisher, J.D. Medaglia, B.F. Jeronimus, Lack of group-to-individual generalizability is a threat to human subjects research, Proceedings of the National Academy of Sciences of the United States of America. 115 (2018) E6106–E6115. https://doi.org/10.1073/pnas.1711978115.

[7] A.M. Beltz, A.G.C. Wright, B.N. Sprague, P.C.M. Molenaar, Bridging the Nomothetic and Idiographic Approaches to the Analysis of Clinical Data, Assessment. 23 (2016) 447–458. https://doi.org/10.1177/1073191116648209.

[8] P.C.M. Molenaar, C.G. Campbell, The New Person-Specific Paradigm in Psychology, Current Directions in Psychological Science. 18 (2009) 112–117. https://doi.org/10.1111/j.1467-8721.2009.01619.x.

[9] J.T. Lamiell, Toward an idiothetic psychology of personality, American Psychologist. 36 (1981) 276–289. https://doi.org/10.1037/0003-066X.36.3.276.

[10] S. Epskamp, L.J. Waldorp, R. Mõttus, D. Borsboom, The Gaussian Graphical Model in Cross-Sectional and Time-Series Data, Multivariate Behavioral Research. 53 (2018) 453–480. https://doi.org/10.1080/00273171.2018.1454823.

[11] G. Costantini, S. Epskamp, D. Borsboom, M. Perugini, R. Mõttus, L.J. Waldorp, A.O.J. Cramer, State of the aRt personality research: A tutorial on network analysis of personality data in R, Journal of Research in Personality. 54 (2015) 13–29. https://doi.org/10.1016/j.jrp.2014.07.003.

[12] M. Saqr, O. Viberg, W. Peeters, Using Psychological Networks to Reveal the Interplay between Foreign Language Students' Self-Regulated Learning Tactics, in: Proceedings of the 2020 STELLA Symposium, 2021: pp. 12–23.

[13] M. Dado, D. Bodemer, A review of methodological applications of social network analysis in computer-supported collaborative learning, Educational Research Review. 22 (2017) 159–180. https://doi.org/10.1016/j.edurev.2017.08.005.

[14] M. Saqr, J. Nouri, U. Fors, Time to focus on the temporal dimension of learning: a learning analytics study of the temporal patterns of students' interactions and self-regulation, International Journal of Technology Enhanced Learning. 11 (2019) 398. https://doi.org/10.1504/IJTEL.2019.102549.

[15] B. Chen, O. Poquet, Socio-temporal dynamics in peer interaction events, in: Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, ACM, New York, NY, USA, 2020: pp. 203–208. https://doi.org/10.1145/3375462.3375535.

[16] B. Chen, T. Huang, It is about timing: Network prestige in asynchronous online discussions, Journal of Computer Assisted Learning. 35 (2019) 503–515. https://doi.org/10.1111/jcal.12355.

[17] I. Halatchliyski, T. Hecking, T. Göhnert, H.U. Hoppe, Analyzing the Flow of Ideas and Profiles of Contributors in an Open Learning Community, Proceedings of the Third International Conference on Learning Analytics and Knowledge - LAK '13. 1 (2013) 66–74. https://doi.org/10.1145/2460296.2460311.

[18] D. Hevey, Network analysis: A brief overview and tutorial, Health Psychology and Behavioral Medicine. 6 (2018) 301–328. https://doi.org/10.1080/21642850.2018.1521283.

[19] R. Artner, P.P. Wellingerhof, G. Lafit, T. Loossens, W. Vanpaemel, F. Tuerlinckx, The shape of partial correlation matrices, Communications in Statistics - Theory and Methods. 0 (2020) 1–18. https://doi.org/10.1080/03610926.2020.1811338.
[20] M. Hamilton, J. Clarke-Midura, J.F. Shumway, V.R. Lee, An Emerging Technology Report on Computational Toys in Early Childhood, Technology, Knowledge and Learning. (2019). https://doi.org/10.1007/s10758-019-09423-8.
[21] D. Borsboom, A network theory of mental disorders, World Psychiatry. 16 (2017) 5–13. https://doi.org/10.1002/wps.20375.
[22] P.C.M. Molenaar, A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever, Measurement: Interdisciplinary Research & Perspective. 2 (2004) 201–218. https://doi.org/10.1207/s15366359mea0204_1.
[23] D. Tempelaar, B. Rienties, J. Mittelmeier, Q. Nguyen, Student profiling in a dispositional learning analytics application using formative assessment, Computers in Human Behavior. 78 (2018) 408–420. https://doi.org/10.1016/j.chb.2017.08.010.
[24] D. Tempelaar, B. Rienties, Q. Nguyen, Investigating learning strategies in a dispositional learning analytics context: The case of worked examples, ACM International Conference Proceeding Series. (2018) 201–205. https://doi.org/10.1145/3170358.3170385.
[25] D. Tempelaar, How Dispositional learning analytics helps understanding the worked-example principle, in: Proceedings 14th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2017), 2017: pp. 117–124.
[26] S. Epskamp, C.D. van Borkulo, D.C. van der Veen, M.N. Servaas, A.M. Isvoranu, H. Riese, A.O.J. Cramer, Personalized Network Modeling in Psychopathology: The Importance of Contemporaneous and Temporal Connections, Clinical Psychological Science. 6 (2018) 416–427. https://doi.org/10.1177/2167702617744325.
[27] A.J. Fisher, J.W. Reeves, G. Lawyer, J.D. Medaglia, J.A. Rubel, Exploring the idiographic dynamics of mood and anxiety via network analysis, Journal of Abnormal Psychology. 126 (2017) 1044–1056. https://doi.org/10.1037/abn0000311.
[28] M. Saqr, Big data and the emerging ethical challenges., International Journal of Health Sciences. 11 (2017) 1–2.
[29] A. Munoz-Arcentales, S. López-Pernas, A. Pozo, Á. Alonso, J. Salvachúa, G. Huecas, An Architecture for Providing Data Usage and Access Control in Data Sharing Ecosystems, Proceedings of the 6th International Symposium on Emerging Information, Communication and Networks (EICN 2019). 160 (2019) 590–597. https://doi.org/10.1016/j.procs.2019.11.042.
[30] S. López-Pernas, M. Saqr, Idiographic Learning Analytics: A Within-Person Ethical Perspective, in: Companion Proceedings 11th International Conference on Learning Analytics & Knowledge (LAK21), 2021: pp. 310–315.