From childhood to maturity: Are we there yet?

Mapping the intellectual progress in learning analytics during the past decade

Zacharoula Papamitsiou
Norwegian University of Science and Technology
Trondheim, Norway
zacharoula.papamitsiou@ntnu.no

Michail N. Giannakos
Norwegian University of Science and Technology
Trondheim, Norway
michailg@ntnu.no

Xavier Ochoa
New York University
New York, USA
xavier.ochoa@nyu.edu

ABSTRACT

This study aims to identify the conceptual structure and the thematic progress in Learning Analytics (evolution) and to elaborate on backbone/emerging topics in the field (maturity) from 2011 to September 2019. To address this objective, this paper employs hierarchical clustering, strategic diagrams and network analysis to construct the intellectual map of the Learning Analytics community and to visualize the thematic landscape in this field, using co-word analysis. Overall, a total of 459 papers from the proceedings of the Learning Analytics and Knowledge (LAK) conference and 168 articles published in the Journal of Learning Analytics (JLA), and the respective 3092 author-assigned keywords and 4051 machine-extracted key-phrases, were included in the analyses. The results indicate that the community has significantly focused in areas like Massive Open Online Courses and visualizations; Learning Management Systems, assessment and self-regulated learning are also basic topics, yet topics like natural language processing and orchestration are emerging. The analysis highlights the shift of the research interest throughout the past decade, and the rise of new topics, comprising evidence that the field is expanding. Limitations of the approach and future work plans conclude the paper.

CCS CONCEPTS

• **Applied computing → E-learning**;

KEYWORDS

Co-word analysis; bibliometrics; conceptual evolution; learning analytics

ACM Reference Format:
Zacharoula Papamitsiou, Michail N. Giannakos, and Xavier Ochoa. 2020. From childhood to maturity: Are we there yet?: Mapping the intellectual progress in learning analytics during the past decade. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Learning analytics is a multidisciplinary, rapidly growing field of research and practice that integrates learning, data sciences and educational technology into a rich socio-technical ecosystem [29]. Given the interdisciplinary nature of Learning Analytics, as well as its novelty and its proximity to other research domains (such as Educational Data Mining (EDM) and Artificial Intelligence in Education), the field, from its early years, has raised questions about itself. These self-reflections of the learning analytics community have been shaped into concrete research questions, addressed in applied bibliometrics analysis or systematic reviews that have explored the structure of the community [29], which research questions and methodologies are more commonly used [8], the intersection of learning analytics with other relevant research fields [30, 34], the impact on our understanding of learning and the contribution to mainstream practice or extending theory [9], and existing evidence on the adoption of learning analytics in higher education [37].

Now that the field is officially ten years old, counting by the number of Learning Analytics and Knowledge (LAK) conferences being organized, is a good timing to analyze its past and current state. An appropriate way to conduct this analysis is applying the same quantitative methodology that is central to the field: this type of analysis is not a navel-gazing exercise, but it is essential activity to quantify the core topics, the marginal contributions, the under-developed themes, and the forthcoming ideas that worth investing on, as well as how these topics related between themselves and move between these states during the last 10 years. The main objective of this work is to produce a *classification schema* of scientific publications that defines the field's own sub-areas that are mostly accepted by the community, which, in turn, consults and feeds them.

Towards facilitating this objective, this paper employs co-word analysis, shows the internal dynamics and structure of the domain, and identifies the topics with impact in the given discipline to date [3]. Co-word analysis allows for and supports the identification of key patterns and trends that point to particular changes in research topics (e.g., emerging or declining interests) or specific research directions (e.g., paradigm shifts), using a graph key-terms [19], *extracted directly from the metadata of the papers.*

Considering this, the present study maps the intellectual progress of the learning analytics landscape, as reflected in the records of flagship publication venues of the Society for Learning Analytics Research (SoLAR), i.e., LAK conference and Journal of Learning Analytics (JLA), which provide a solid foundation to the related work published to date. During the past decade, considerable work has been published, allowing us to observe where the field currently stands, what are the challenges and opportunities the researchers

are facing, and what are the potential driving forces in the near future. Accordingly, this work mainly contributes as follows:

- brings new insights on the intellectual mapping and the evolution of the scientific area of learning analytics;
- raises awareness of the community on the mature, underdeveloped, emerging, or declining research themes;
- highlights individual topics as popular, core or backbone research topics within the discipline.

2 BACKGROUND AND RELATED WORK

The community of learning analytics is rapidly expanding [28, 29]. The quantity and quality of the research activity within this community has been the topic to a variety of bibliometrics and literature reviews, aiming to evaluate the research progress, impact and societal value, from different viewpoints (e.g., [8, 9, 38]).

Specifically, early work emphasized on the dimensions of learning analytics [4] as well as the drivers, developments, and challenges [13]. Papamitsiou and Economides [30] reviewed the empirical evidence of learning analytics and Educational Data Mining (EDM), and suggested a classification of previous research works according to numerous criteria (e.g., objectives, methods, context), in a systematic manner. The identified key topics included student modeling, prediction of performance, dropout and retention, increase of reflection and awareness, improvement of feedback/assessment services, and recommendation of resources. More recently, the same objective, yet from a different perspective, was addressed by Dawson et al. [9]. The authors coded the papers according to five dimensions — Focus, Purpose, Scale, Data and Settings – and applied epistemic network analysis. They came to the conclusion that while there is substantial research in the areas of focus and sophistication of analyses, the focus on practice, theory and frameworks is limited.

A comprehensive overview of the evolution of learning analytics from a pedagogical perspective was presented in a state-of-the-art report [24], and the current application of learning analytics in Higher Education (HE) with a focus on research approaches, methods, and evidence was also reviewed [37]. The analysis indicated that, although the learning analytics field is maturing and shifting towards a deeper understanding of students' learning experiences, the overall potential of learning analytics in HE is far from adoption.

Complementary to the qualitative literature review studies, bibliometrics is a commonly employed quantitative approach, introduced as a statistical method that uses various indicators (e.g., citations, authors, number of publications) to examine the performance and development of a field [31]. There are several ways to analyze and map bibliometrics data: co-author analysis is adopted to study the social structure within a particular field [17], co-word analysis studies the conceptual structure of a research field [22], and co-citation analysis aims to identify which papers and authors are most referenced, and how highly-cited papers are connected [35].

In the field of learning analytics, the identified topics from the papers in LAK2013 conference were visualization, behaviour analysis, social learning analytics, learning analytics for MOOCs, and learning analytics issues (e.g., ethical, scalability, etc.) [29]. In an analogous approach, a co-author and citation analysis mapped the research community, and identified the emergence of trends and disciplinary hierarchies, and the diversity of research genres [8]. The analysis detected some fragmentation in the major disciplines (i.e., computer science and education) regarding conference and journal representation. Recently, Waheed et al. [38] analyzed publication counts, citation counts, co-authorship patterns, citation networks, and term co-occurrence, and identified that the terms "students", "teachers", "higher education institutions", and "learning process" appear to be the major components of the field so far.

As learning analytics is a continuously evolving field, it is important to (a) identify and understand its core foundations that might contribute to reinforcing the community's identity; (b) detect underrepresented or under-developed themes that require attention for their inclusion and success; (c) highlight research gaps in bridging theory and practice; and (d) find challenges and opportunities that hold the promise for improving the educational processes.

3 METHODOLOGY

3.1 Data collection

The data analyzed in this study were downloaded from the ACM Digital Library (i.e., papers published in LAK proceedings between 2011 and 2019) and from the Journal of Learning Analytics (i.e., articles published between 2014 and September 2019). Overall, 627 peer-reviewed articles (full and short LAK papers, and JLA papers) have been published to-date. Among them, 168 were published in JLA and the rest 459 papers were published in LAK. The editorials from JLA were excluded from the analysis. From the collected papers, the author-assigned keywords were extracted from the metadata of each paper and were used as a unit of analysis. However, those keywords can be potentially biased to human subjectivity; for example, the authors might use more generic terms to describe their work to ensure its visibility, to categorize and link their work to a broader research domain or to synopsize the sub-topics and replace specific terms with more generic ones (e.g., "machine learning" instead of "Random Forest, Neural Networks"). Therefore, the abstracts of the papers were also text-mined in order to automatically extract from them key-phrases that can describe their contents, based on the "agreement" that the abstract can be seen as a "stand-alone" version of the paper, as it synopsizes the paper in a coherent manner. The 627 papers, containing 3092 author-assigned keywords (M=4.93 per article) and 4051 (M=6.46 per article) machine-extracted key-phrases, are distributed per year of publication as shown in Figure 1.

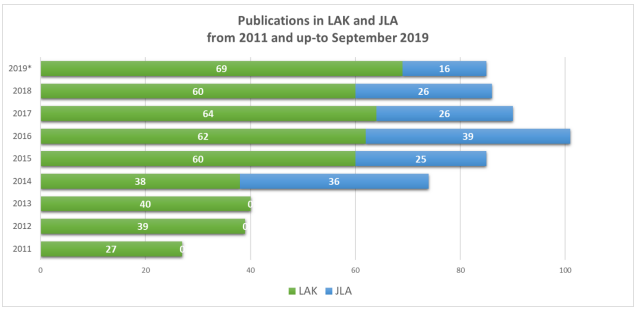

Figure 1: Number of Learning Analytics publications (LAK-JLA) per year for the period 2011-2019 (Sept.)

3.2 Data pre-processing

The retrieved author-assigned keywords were manually pre-processed and standardized through merging words appearing in singular

and plural forms of nouns, words that convey similar meaning (e.g., "information visualization" and "data visualization" were merged into "visualization"), fixing misspelled keywords (e.g., "leaning analytics"), following a common spelling for UK and US terms (e.g., "behaviour" and "behavior"), and filtering broadly used terms (e.g., "SNA" and "Social Network Analysis"; "MOOCs" and "Massive Open Online Courses") - following the approach recommended in [19, 22, 39], in a non-invasive manner. At the end of this phase, 1581 (51.1% of the originally author-assigned keywords) were identified as *unique* keywords, and were subjected to further analysis.

For extracting the key-phrases from abstracts, an implementation of the TextRank algorithm in Python for text summarization was used [23]. TextRank is fully unsupervised, i.e., no training is necessary, and instead of n-grams, it can tokenize and annotate with Part of Speech (PoS). In this study, the TextRank sliding window was set to 3, for the PoS we included nouns (NOUN), adjectives (ADJ) and proper nouns (PROPN), and we requested for the top-10 phrases. Since not all phrases are highly semantic (e.g., "general goal", "first iteration", "contribution"), after manually removing those phrases, we retained a total of 4051 key-phrases, and we repeated the same pre-processing as for the author-assigned keywords, ending-up with 2525 (62.3%) key-phrases identified as *unique*.

The Kolmogorov Smirnoff test shown that the frequency of keywords in both cases follows a power-law distribution with an alpha of 1.95 and 2.03 respectively. Due to this heavy-tailedness, the research landscape of learning analytics is a *scale-free network*, with small number of popular terms acting as *"hubs"*: they connect different topics, capture major research directions and major influences in the field, and shape the intellectual structure of the field [19, 39].

A scale-free network also suggests that major research themes can be detected with small subset of popular terms [21]. A previous analysis in the HCI research field demonstrated that less than 100 keywords are enough to describe the intellectual progress of a field [22]. Thus, in the present study we decided to include only those key-terms that appear more than six times (n≥ 6) in the period 2011-2019. This decision was grounded on two facets: (a) the frequency of a term reflects its significance for a research community, i.e., the higher the frequency is, the more often the term attracts the researchers' attention/interest; and (b) the retained 59 authors-assigned keywords (total frequency=793, 50.2% of the total keywords) cover 553 (88.2%) of the 627 articles published, whereas the 85 machine-extracted key-phrases (total frequency=1094, 43.3% of the total phrases) cover 583 (93%) of the initial articles. Furthermore, for the given datasets of terms and papers, n=6 is the minimum term frequency that achieves the highest inclusion of papers in the datasets. For example, for author-assigned keywords with n≥ 5, the retained keywords are N=66 and cover 88.9% of the papers, whereas for keywords with n≥ 7, N=52 keywords, covering 77.3% of the papers. Similar are the results for the machine-extracted key-phrases. Thus, with fewer yet highly frequent terms we could satisfactorily describe the Learning Analytics network of terms.

3.3 Co-word analysis and strategic diagram

As stated in the introduction, this study employs co-word analysis to understand the big picture of learning analytics research. Co-word analysis has been proposed as a content analysis technique to map the strength of relation between terms in texts and to trace patterns and trends in term associated-ness [3]. The idea behind co-word analysis rests on the assumption that key-terms identified within an article (either as author-assigned keywords or as machine-extracted key-phrases) can adequately describe and communicate the content of that article; the co-occurrence of two (or more) key-terms in the same article indicates a linkage between those topics, i.e., a *"theme"* [2]. The main units of analysis are *key-terms*, *clusters* (i.e., sets of closely-related key-terms) and *key-term networks* [22].

Co-word analysis is applied to reduce the broad network of key-terms into a smaller network of related topics using graph theory [6]. Graphs consist of nodes that represent the key-terms, and links that represent the interactions between the nodes. Given a network of key-terms, a combination of clustering, network analysis and strategic diagrams is used to model the conceptual structure of a field [2]. The graph theory concepts employed to map the research field are *centrality* (i.e., the strength of the links from one research theme or cluster to others, indicating its significance in the development of the community [22]) and *density* (i.e., the coherence of a cluster and a measure of a theme's development [18]). Combining centrality (*x-axis*) and density (*y-axis*) allows for creation of two-dimensional *strategic diagrams* [2]: the position of a cluster in the diagram corresponds to the importance of the cluster in the whole network (i.e., centrality) in relation to how well the theme of this cluster is developed (i.e., density), shown in Figure 2.

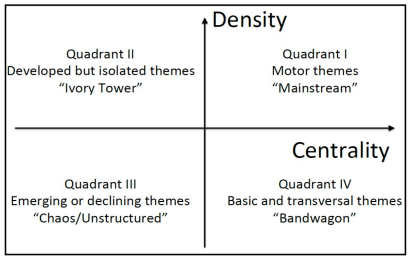

Figure 2: Strategic diagram of density and centrality [22]

As one can observe, *Quadrant I (Q1)* holds the motor themes (i.e., mainstream themes) that have strong centrality and high density. *Quadrant II (Q2)* contains themes that are internally well-structured, but have weak external ties. These research themes are more specialized and peripheral to the mainstream work that is central in the research field. *Quadrant III (Q3)* includes the themes with low density and low centrality, that are either emerging or disappearing. Finally, *Quadrant IV (Q4)* covers basic and transversal themes, central to the community, holding the potential to become significant.

3.4 Data analysis

To identify the major research themes in the learning analytics domain, hierarchical clustering analysis on a correlation matrix with the retained terms was performed, using the Ward's method with Squared Euclidean Distance as the distance measurement [25]. The supervised clustering method allows to maintain content validity and cluster fitness for the highest number of clusters [19, 22]. Each cluster represents a research theme or sub-field. The co-word network was further analyzed using the following measures:

- **Key-terms:** set of terms that constitute a cluster;
- **Size:** number of key-terms in the cluster;

- **Frequency:** how many times all key-terms (in a cluster) appear in the dataset;
- **Co-word frequency:** how many times at-least two key-terms (from a cluster) appear in the same paper. Computing the frequency of two terms appearing together in the same paper results in a symmetrical co-occurrence matrix [20]. In this matrix, values in the diagonal cells are term frequencies, and values in non-diagonal cells are co-word frequencies. High frequency of co-occurrence between terms indicates connection between the topics they represent;
- **Transitivity:** how tightly connected is the cluster (the *clustering coefficient*), i.e., how close the key-terms are to being a "clique". Transitivity is the frequency of loops of length three in the cluster; a loop of length three is a sequence of nodes x, y, z such that (x, y), (y, z) and (z, x) are edges of the graph [33]. The value range for transitivity is [0, 1];
- **Centrality:** the degree of interaction of a theme with other parts of the network, i.e, how many other clusters a cluster connects to [2]; Centrality refers to a group of metrics that aim to quantify the "importance" of a particular node (or cluster) within a network (e.g., betweenness centrality, closeness centrality, eigenvector centrality, degree centrality) [26]. Here we used betweenness centrality (C), with $0 \leq C \leq 1$;
- **Density:** how cohesive is the cluster of terms, i.e, the number of direct ties observed for the cluster divided by the maximum number of possible ones [2]. Density is graph-dependent, and can be any positive real number [10].

Based on the clustering results, we plotted the strategic diagram for the years 2011-2019 to visualize the cohesion and maturity of the research themes in learning analytics [2, 22].

We repeated this approach for the author-assigned keywords and the machine-extracted key-phrases, as well as for the merged list of terms. The "merging" decision was driven by the assumption that the machine-extracted phrases can reduce the bias inserted by human-judgement, whereas, considering the human generalization capability can facilitate a "highly-semantics text-annotation" process for keyword selection. Duplicates of key-terms in the merged list were removed in order not to insert bias in the dataset. It should be made explicitly clear that the selection of key-terms is not a "taxonomy" created by the authors of this paper, but a collection of terms that are extensively found in the learning analytics literature.

In addition, a key-term network graph was created from the key-term list. In this graph, each key-term is represented as a node, and the key-terms that co-appear on a paper are linked together. By creating associations between key-terms, multiple networks associated with different themes are also created. In this case, bridges are built between the nodes of key-terms, to allow communication and information flow between isolated regions in the whole network. Those nodes are known as *structural holes* [27]. Key-terms acting as structural holes also serve as a "backbone" of a network: if removed, the network will loose its cohesion and will disintegrate into separated and unconnected concepts. Thus, the network's core-periphery structure needs to be computed, in order to determine which nodes are part of a densely connected core (i.e., with a higher number of bridges) or a sparsely connected periphery core [32]. Core nodes are reasonably well connected to peripheral nodes, while peripheral nodes are sparingly connected to a core node or to each other. Hence, a node belongs to a core only if it is well-connected to other core nodes and to peripheral nodes [32]. A follow-up core-periphery analysis was performed to spot the core research topics from the perspective of the whole network. In this analysis, key-terms were categorized according to their popularity, coreness (i.e., connectedness with other topics) and constraint (i.e., backbone topics). The whole approach is illustrated in Figure 3.

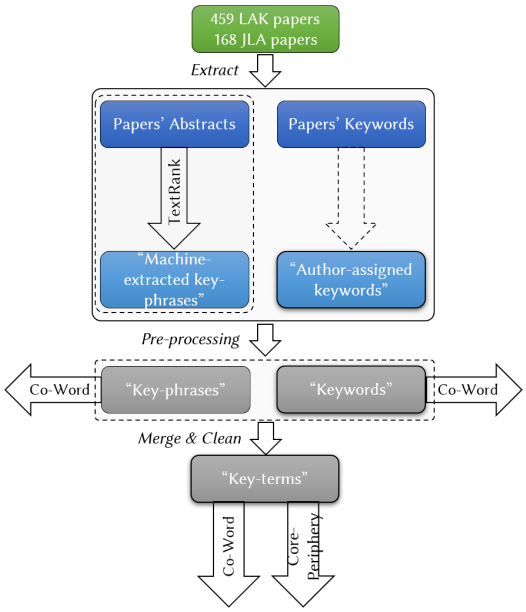

Figure 3: Research Methodology

4 RESULTS

4.1 Mapping of the field: the authors' perspective and the machine intelligence

The analysis on the retained 59 and 85 author-assigned keywords and machine-extracted key-phrases led to 14 clusters in each case (labeled as C1-C14, in Tables 1 and 2 respectively), with each cluster representing a research theme or a sub-field. In order (a) to better understand the relative "position" of these clusters within the overall learning analytics field (i.e., what is the distance from each other in terms of cohesion and maturity of research themes they correspond to); and (b) to create the conceptual structure of the learning analytics discipline, we constructed strategic diagrams (plots) using the centrality and density of each cluster [2, 22]. The overall results can be seen in reading Figure 4 and Table 1, and Figure 5 and Table 2 together, respectively. In the plots, both axes are centralized to the average centrality and average density respectively (i.e., 0.452, 2.955 for author-assigned keywords and 0.584, 1.913 for machine-extracted key-phrases). The overall networks' densities, were 0.195 and 0.243 respectively.

As it can be observed from Figure 4, one motor theme (*Mainstream theme*), represented by cluster C12 (i.e., MOOCs, SNA, discussion forum) is detected when using the human descriptors (keywords) of the papers. In other words, the field is in general fragmented, with only one theme having received substantial attention

from the community, in terms of human annotations. However, although a similar cluster is also identified as motor theme when the machine-extracted phrases are employed, in this case, additional topics are also categorized as mainstream. Specifically, as shown in Figure 5, clusters C5, C11 and C7 (i.e., LMS, engagement, MOOCs, discussion forums, assessment, personalization) are marked as leading the field of learning analytics research and appear to have been well-structured and strongly-tied to the other research topics.

Furthermore, in Figure 4, the author-assigned keywords indicate that the community has few internally well-structured research themes, yet with weak external ties (*Ivory Towers*), acting as peripheral nodes to the global network (i.e., connect only to core nodes, yet not necessarily to mainstream topics only), and classified in clusters C1, C3, C5 (e.g., visualization, social learning analytics, retention). Visualization is also marked as "ivory tower" using the machine-extracted phrases, along with learning design, temporal learning analytics and reflection, representing clusters C6, C4, C9 and C2 respectively in Figure 5. Those topics appear to have high-density, i.e., the clustering coefficient is high and the topics within each of the cluster are very well connected to each other, but they lack strong ties with topics that are external to them. The following-up core-periphery analysis will provide insight on that issue.

Regarding the themes that are either emerging or disappearing (*Chaos/Unstructured*), the author-assigned keywords revealed that researchers have developed a considerable number of topics with – in a sense – "marginal" interest in the learning analytics network, classified in clusters C2, C4, C6, C8 and C13 (e.g., classroom, higher education, LMS, self-regulated learning, CSCL), as illustrated in Figure 4. The term "marginal" here is used to describe both the cases of "close-to-diasappearing" and "nearly rising" topics, i.e., topics that either tend to no-longer attract major interest or they have recently started to attract attention, but have not yet been well-developed. However, the results from the machine-extracted key-phrases for this category of topics compose a significantly different landscape: Figure 5 demonstrates two chaotic or unstructured themes, i.e., C8 and C12 (e.g., SNA, social learning, dataset, ethics).

Finally, a substantial number of *Bandwagon* themes that are central for the research community, yet are only weakly linked *together* (i.e., they comprise sparse graphs of topics) have been detected as well. From the author-assigned keywords (Figure 4), those topics are categorized in clusters C7, C9, C10, C11 and C14 (e.g., NLP, EDM, assessment, ITS, engagement). The respective topics – identified using the terms from the text-mined abstracts – are included in C2, C3, C10, C13, C14 in Figure 5 (e.g., classroom, behaviour, performance, higher education, classification).

The differences in the results between the human selected descriptors and the automatically extracted ones, motivated us to also consider the merged dataset of key-terms for further analysis.

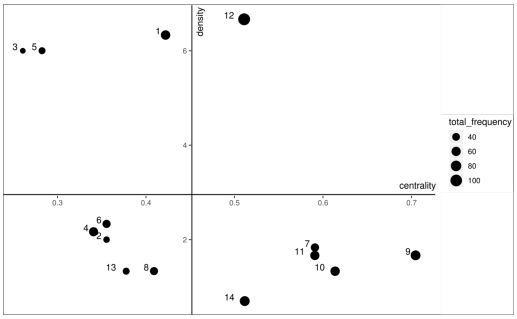

Figure 4: Strategic diagram for learning analytics for the period 2011-2019 - Author-assigned keywords

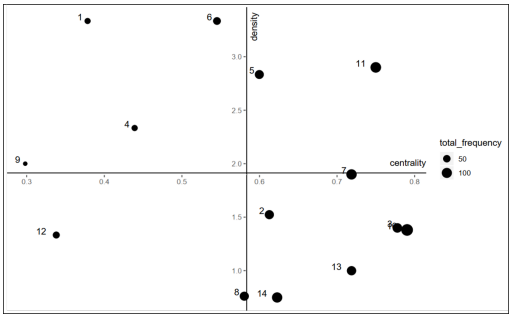

Figure 5: Strategic diagram for learning analytics for the period 2011-2019 - Machine-extracted key-phrases

4.2 Enhancing human judgement with insights from the machine

Table 1: Clusters of topics in Learning Analytics for the period 2011-2019 - Author-assigned keywords

Q	ID	Key-terms (the most frequent being in bold)	Size	Freq[1]	CW-Fr.[1]	T[1]	C[1]	D[1]
Q1	C12	**MOOCs**, SNA, discussion forum	3	105	104	1.00	0.51	6.67
Q2	C1	**visualization**, awareness, dashboards	3	63	69	1.00	0.42	6.33
Q2	C3	discourse analytics, **social learning analytics**, social learning	3	26	29	1.00	0.26	6.00
Q2	C5	predictive analytics, **retention**, student success, early intervention	4	41	48	1.00	0.28	6.00
Q3	C2	orchestration, **classroom**, teachers, co-design, k12	5	36	44	0.78	0.36	2.00
Q3	C4	distance education, **higher education**, policy, ethics	4	58	56	0.83	0.34	2.17
Q3	C6	**LMS**, data mining, early warning systems	3	45	38	1.00	0.36	2.33
Q3	C8	**self-regulated learning**, learning, analytics, metacognition	4	44	49	0.83	0.41	1.33
Q3	C13	**CSCL**, measurement, eye tracking, 21st century skills	4	36	39	0.90	0.38	1.33
Q4	C7	education, reflection, **NLP**, topic modeling, corpus linguistics, stealth assessment, writing	6	63	74	0.52	0.59	1.83
Q4	C9	**EDM**, learner modeling, blended learning, clustering	4	69	79	1.00	0.70	1.67
Q4	C10	**assessment**, feedback, MMLA, personalization	4	63	61	0.83	0.61	1.33
Q4	C11	**ITS**, predictive modeling, mathematics, machine learning, bayesian knowledge tracing	5	65	83	0.90	0.59	1.67
Q4	C14	learning design, online learning, collaborative learning, **engagement**, classification, survival analysis, temporal analysis	7	79	75	0.64	0.51	0.70

[1] **Freq:** Total frequency of all key-terms; **CW-Fr:** Co-word Frequency; **T:** Transitivity; **C:** Centrality; **D:** Density

Table 2: Clusters of topics in Learning Analytics for the period 2011-2019 - Machine-extracted key-phrases

Q	ID	Key-terms (the most frequent being in bold)	Size	Freq[1]	CW-Fr[1]	T[1]	C[1]	D[1]
Q1	C5	grades, **LMS**, engagement, course material, resource usage	5	82	148	1.00	0.60	2.83
Q1	C11	**MOOCs**, content analysis, discussion forums, NLP, topic modeling, linguistic features	6	115	197	0.90	0.75	2.90
Q1/Q4	C7	**assessment**, feedback, mathematics, personalization, achievement	5	105	179	0.83	0.72	1.90
Q2	C1	self-reports, self-efficacy, **reflection**, self-explanation	4	21	83	1.00	0.38	3.33
Q2	C4	**learning design**, eye-tacking, multimodal learning analytics, trace data, attention, actionable insights	6	45	64	0.88	0.44	2.33
Q2	C6	dashboards, **visualization**, self-regualted learning, visual analytics, metacognition	5	68	109	1.00	0.55	3.33
Q2	C9	**temporal learning analytics**, time-series, sequences	3	14	34	1.00	0.29	2.00
Q3	C8	**SNA**, informal learning, social learning, **collaborative learning**, 21st century skills, participation, virtual learning environments, social media, social learning analytics	9	87	122	0.62	0.58	0.76
Q3	C12	big data, ethics, privacy, **dataset**, policy	5	47	53	0.78	0.34	1.33
Q4	C2	CSCL, collaboration, **classroom**, real-time, design, awareness, orchestration, design process, HCI	9	84	164	0.87	0.61	1.52
Q4	C3	learning resources, learning processes, **behaviour**, online learning environments, log data, clickstreams	6	88	150	0.78	0.78	1.40
Q4	C10	learner models, **performance**, learning outcomes, ITS, prediction, analytics, discourse analysis, prior knowledge, learning gains	9	151	218	0.83	0.79	1.38
Q4	C13	predictive modeling, **higher education**, academic performance, dropout, programming	5	83	136	0.78	0.72	1.00
Q4	C14	EDM, student success, data mining, retention, early warning systems, machine learning, **classification**, clustering	8	104	167	0.90	0.62	0.75

[1] **Freq:** Total frequency of all key-terms; **CW-Fr:** Co-word Frequency; **T:** Transitivity; **C:** Centrality; **D:** Density

4.2.1 Correspondence analysis. The merging of the author-assigned keywords with the machine-extracted key-phrases yielded a list of 84 key-terms (that appear more than 6 times, as previously, after removing the duplicates), covering 93.6% of the 627 papers. To gain a first understanding and insight from those key-terms, correspondence analysis (CA) was applied. CA is a descriptive, exploratory technique suited to handle categorical data, both graphically and numerically. When applied on a dataset of words, the standard coordinates show the position of the words on the underlying dimensions (i.e., factors). The results are interpreted based on the relative positions of the points and their distribution along the dimensions; as words are more similar in distribution, the closer they are represented in the map [7]. In other words, CA is employed to to grasp the overall topics distribution and how they are spread from 2011 to 2019 based on the occurrences (frequencies) of the 84 key-terms throughout the years. To achieve this, CA performs a homogeneity analysis of a contingency table to obtain a low-dimensional Euclidean representation of the original data [16]. CA is used to analyze frequencies formed by categorical data (i.e, contingency table) and it provides factor scores (coordinates) for both

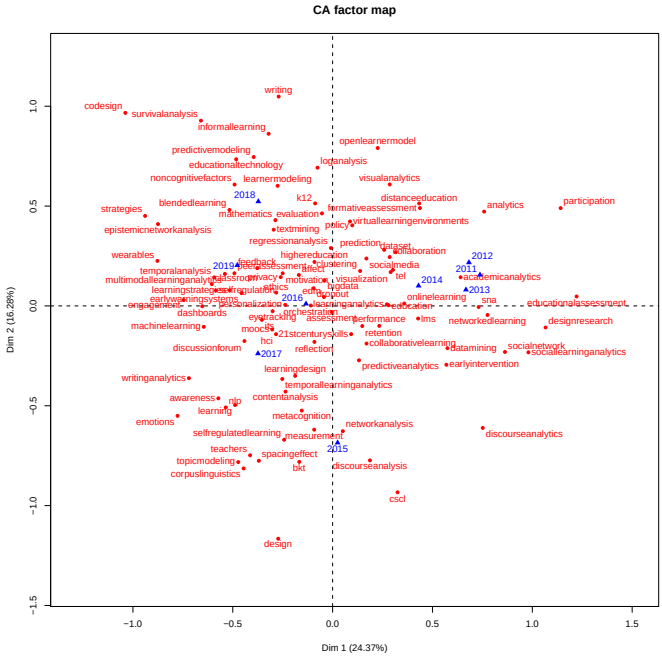

Figure 6: Correspondence Analysis map for Learning Analytics for the period 2011-2019

the rows and the columns of contingency table. These coordinates are used to visualize graphically the association between the row and column variables in the contingency table in a two-dimensional space, based on the chi-squared statistic associated with the contingency table. In the two-dimensional outcome chart, all rows and all columns of the contingency table can be displayed on the same axes. The results of the correspondence analysis for Learning Analytics for the years 2011-2019 are illustrated in Figure 6. The rows and columns of the contingency table are the years and key-terms. The percentages on the axes correspond to the variance explained by the two dimensions considered together. Summing up the proportion of variance explained by both dimensions shows how much of the variance in the data can be explained by the visualization. In this study, the visualization displays 40.65% of the variance in the data.

The CA factor map positions all key-terms and years on a common set of orthogonal axes, illustrating which terms are most frequently met on a specific year. For example, as seen in Figure 6, "'academic analytics" was more frequently used in 2011, "online learning" was more regularly found in 2013 papers, "orchestration" was appearing more times in 2016, whereas "noncognitive factors, K12" are key-terms that were more likely to be found in 2018 literature. It seems that there was a significant jump in the diversity of Learning Analytics topics happening in 2015. The position of each year after 2015 appears to move more aggressively compared to the "smoother" (less diverse) shift in topics during the years 2011-2014, indicating a sharp diversification of topics from that time point on.

4.2.2 Co-word analysis. The 84 retained key-terms shaped 14 clusters of themes, and their relative positions in terms of centrality and density in the learning analytics term-network are illustrated in Figure 7. The full description of the clusters is provided in Table 3. For this network of terms, the average centrality is 0.757, the average density is 3.238 and the overall density is 0.175.

From Figure 7 it can be observed that within the merged body of key-terms there are two motor themes for the learning analytics filed, represented by clusters C2 and C11 (i.e., visualization, MOOCs, discussion forum, awareness). Again, some degree of fragmentation in the field is revealed, indicating that the community has paid substantial attention mostly on those topics. In addition, only two themes are identified as peripheral to the whole network. The ivory towers in this case are the topics classified in C3 and C13 (e.g., ethics, social learning analytics). On the contrary, the chaotic/unstructured topics that have started either to emerge or to disappear, seem to constitute a substantial body in the learning analytics literature. Those themes are clustered in C1, C4, C6, C7 and C9 and comprise e.g., dropout, participation, classroom, orchestration, eye-tracking and NLP (detailed list can be found in Table 3). Finally, a considerable number of Bandwagon themes, strongly linked to specific research interests throughout the network, yet only weakly linked *together* are categorized in clusters C5, C8, C10, C12 and C14 (e.g., LMS, self-regulated learning, engagement, learner modeling, personalization, assessment, learning design).

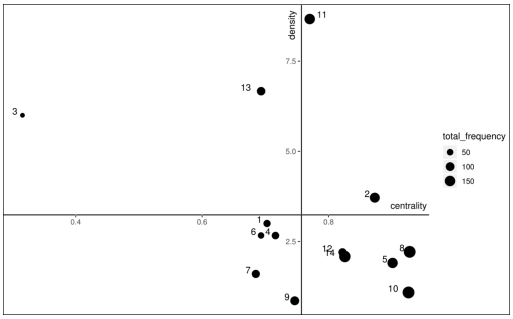

Figure 7: Strategic diagram for learning analytics for the period 2011-2019 - Merged keyterms

4.2.3 Key-terms network map. Overall, a network of key-term demonstrates the relationships among different research themes. To better understand and visualize the interactions between the research themes presented in Table 3, network analysis was performed to create a granular network map of the key-terms. Figure 8 displays these results, in which each node in the graph represents a key-term that is linked to other key-terms that appear on the same

Table 3: Clusters of topics in Learning Analytics for the period 2011-2019 - Merged key-terms

Q	ID	Key-terms (the most frequent being in bold)	Size	Freq[1]	CW-Fr.[1]	T[1]	C[1]	D[1]
Q1	C2	**dashboards**, **visualization**, HCI, collaboration, design, awareness, design process	7	134	419	0.90	0.87	3.71
Q1	C11	**MOOCs**, content analysis, SNA, discussion forum	4	150	334	1.00	0.77	8.67
Q2	C3	discourse analytics, **social learning analytics**, social learning	3	25	49	1.00	0.32	6.00
Q2	C13	ethics, privacy, **higher education**, survey	4	111	200	1.00	0.69	6.67
Q3	C1	teachers, **classroom**, orchestration, real time	4	64	202	1.00	0.70	3.00
Q3	C4	**collaborative learning**, eye-tracking, collaboration, 21st century skills, CSCL	5	78	182	0.90	0.72	2.67
Q3	C6	**education**, reflection, NLP, reflecting writing, linguistics	5	56	132	0.83	0.69	2.67
Q3	C7	learning, analytics, **participation**, online learning, informal learning	5	85	172	0.77	0.68	1.60
Q3	C9	big data, **behaviour**, dataset, dropout, regression analysis, temporal analysis, clickstreams	7	104	225	0.83	0.75	0.86
Q4	C5	formative assessment, **assessment**, feedback, programming, writing, learning sciences, visual analytics	7	145	378	0.87	0.90	1.91
Q4	C8	**EDM**, learner modeling, ITS, learning outcomes, prediction, multimodal learning analytics, machine learning, mathematics	8	196	485	0.62	0.93	2.21
Q4	C10	**self-regulated learning**, learning strategies, learning processes, motivation, personalization, performance, academic performance, blended learning, logdata, clustering	10	198	479	0.66	0.93	1.09
Q4	C12	learning design, **engagement**,trace data, attention, virtual learning environments, learning processes	6	92	249	0.78	0.82	2.20
Q4	C14	grades, predictive analytics, **LMS**, predictive modeling, retention, early warning systems, student success, data mining, classification	9	186	482	0.72	0.83	2.08

[1] **Freq:** Total frequency of all key-terms; **CW-Fr:** Co-word Frequency; **T:** Transitivity; **C:** Centrality; **D:** Density

paper. The size of the nodes is proportional to the frequency of the key-terms, the color of the node corresponds to the cluster the key-term has been classified in, and the thickness of the links between the nodes is proportional tho the co-occurrence correlation for that pair of key-terms. From this analysis, key-terms with less than 5 strong ties, i.e., weak ties, that would lead to a highly disconnected network. To reduce visual clutter, Figure 8 illustrates a centralized subset of the complete network, omitting isolated nodes.

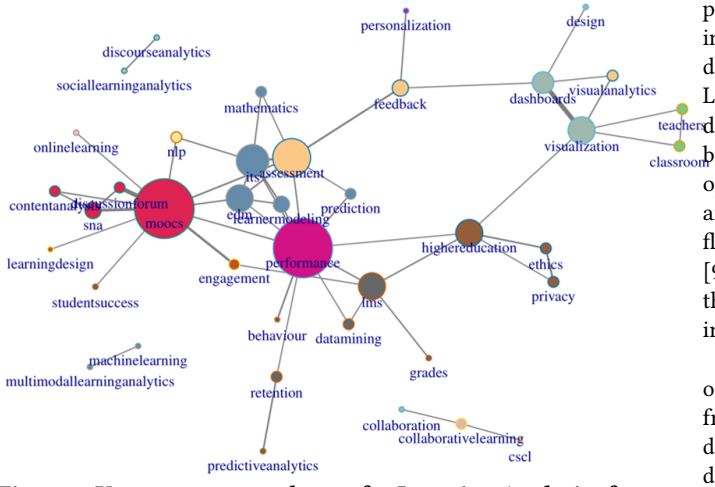

Figure 8: Key-terms network map for Learning Analytics for the period 2011-2019

Finally, core-periphery analysis was performed to identify the core research topics in the field, from a whole-network perspective, as individual key-terms, regardless of the cluster they belong to. The analysis yielded 12 research topics in each of the following categories (Table 4):

- **Popularity**: how frequently a key-term is used;
- **Coreness**: how connected is a key-term with other topics; coreness is measured on a [0-1] scale;
- **Constraint**: how connected is a key-term with other *otherwise distinct* topics (i.e., if the topic creates a backbone of the field); constraint is measured on a [0-1] scale.

High core value indicates a topic that is well connected to other topics. Higher structural holes suggest a key-term that brings together otherwise isolated topics. Burts constraint (i.e., Constraint) [1] is commonly used as a measure of structural holes (accurately speaking, the lack of it, because the larger the constraint value, the less structural opportunities a node may have for bridging structural holes). In other words, key-terms that "act" as bridges between topics have lower constraint values. Topics with high scores on Popularity and Coreness and low score on Constraint can be considered as the driving force for advancements in the field: without these topics, the field of learning analytics would be fragmented.

5 DISCUSSION AND CONCLUSIONS

The field of learning analytics is rapidly expanding as a multidisciplinary domain [28]. In search of its identity through the topics of interest, as they are reflected on the key-terms included in the metadata of the papers published in the two SoLAR flagship venues (i.e., LAK and JLA), from 2011 to 2019 (September), there is a need to understand where we are up-to-date. Previous literature reviews and bibliometrics (e.g., [28–30]) attempted to shed light to the trends, opportunities, and potential challenges in the learning analytics area, as well as to evaluate how much learning analytics have influenced theory and practice and how close we are to adoption [9, 37]. The intent of this study is not to script the field or quantify the impact of previous research to-date, but to raise awareness, influence and support the community in its further maturation.

To address this objective, this study applied co-word analysis on terms from the learning analytics literature that were extracted from the metadata of research work published during the past decade. Specifically, we applied co-word analysis on three different datasets of term that reflect the content of the papers: (a) the author-assigned keywords, which communicate the human perspective; (b) the machine-extracted key-phrases, that mine the significant information directly from text; and (c) the merged key-terms, that enhance the human judgement with insights from the machine. The key findings are critically discussed in relation to previous work.

5.1 Phase 1: Development and Trends

During the last decade the learning analytics community has witnessed steady growth, as evidenced from the increasing number of publications each year (Figure 1). This expands our community to new areas (e.g., classroom, reflective writing) but also helps to develop mainstream/established areas (e.g., self-regulated learning, MOOCs, SNA). The intellectual progress of the field is reflected on the distribution of key-terms throughout the years. As seen from Figure 6, correspondence analysis revealed that, after the first 4 years, learning analytics as a research domain is continuously progressing and shifting towards new research topics, shaping a highly diverse landscape. For example, from "academic analytics" that was a common topic in 2011, the community has gradually move to

Table 4: Summary of popular, core and backbone topics of Learning Analytics in 2011-2019.

#	Popular Topic	Frequency	Core Topic	Coreness [0-1]	Backbone Topic	Constraint [0-1]
1	**MOOCs**	78	**performance**	0.119	**MOOCs**	0.180
2	**performance**	63	**MOOCs**	0.116	**performance**	0.187
3	**higher education**	57	**assessment**	0.116	**assessment**	0.284
4	**assessment**	54	**higher education**	0.114	**LMS**	0.286
5	visualization	50	**EDM**	0.111	**feedback**	0.346
6	**ITS**	39	**LMS**	0.109	**ITS**	0.348
7	**EDM**	37	**ITS**	0.108	**higher education**	0.363
8	**feedback**	37	**feedback**	0.107	**EDM**	0.426
9	**LMS**	34	prediction	0.102	visualization	0.429
10	SNA	34	learner modeling	0.100	collaborative learning	0.500
11	behaviour	33	engagement	0.100	NLP	0.500
12	engagement	33	data mining	0.096	dashboards	0.505

"personalization" is 2016 and has recently shifted the focus on other topics like "noncognitive factors" in 2018 and "peer assessment" in 2019. This spreading of topics is also reflected in the low overall density of the key-terms network (0.175): the multidisciplinary interests are mapped to a sparse network of key-terms, covering multiple aspects of the field and bridging isolated research areas.

The overall results presented in Figure 7 and Table 3 together showcase that the community has developed sufficiently some topics. Two motor themes, clustered in C2 and C11 (i.e., MOOCs and visualizations) appear to be leading the field, being internally well-structured and having strong ties with the other clusters and the external to them network of topics. In addition, those topics are also characterized as popular and backbone (Table 4) throughout the whole network of key-terms, indicating a degree of maturation in research in the respective areas. In a sense, this finding is not surprising since from the early literature reviews and bibliometrics, those topics had gained increased attention [29, 30], and until today, scholars are still working on those sub-areas, as it was recently confirmed that there is substantial research in the areas of focus [9]. We should not forget that some of the mostly cited publications in our field are discussing issues relevant to these themes (e.g., [5, 12]).

Furthermore, the area has also peripheral topics that are well-structured but present limited connections to other topics in the whole network. Indeed, as seen in Figure 8, the topics from cluster C3 (i.e., social learning analytics) are isolated in terms of links to other topics. However, the topic itself appears to be well-established as a stand-alone theme. At the same time, the discussion about ethics and privacy in learning analytics is not new [14, 29], and is also well grounded in theoretical frameworks [11, 36], however, those frameworks are still external to the core of the field itself.

One of the most intriguing findings is that the area has numerous emerging or disappearing topics. As seen from the CA map (Figure 6), some topics have only recently attracted research interest (e.g., multimodal learning analytics), implying that the community follows-up with recent technological advancements, but those topics did not have yet the time required to become central or well-structured (in Q3 in Figure 7). Furthermore, some other topics are gradually fading-out, making space for new ones to rise: this happens either because they have been sufficiently studied (e.g., dropout, clickstreams), or because they are gradually replaced or absorbed by broader terms (e.g., participation with engagement).

One can notice that some topics in the strategic diagram in Figure 7, i.e., for the merged dataset, appear to have moved from one quadrant to another, compared to the strategic diagrams in Figures 4 and 5. A reason why this might happen is the frequency of particular terms which authors use in their works. For example, MOOCs have been a research topic since the very first years of the research discipline and the term "MOOCs" has a shared meaning across the educational research community. On the other hand, visualizations is also a topic that has been extensively studied since the early years of learning analytics. However, authors use either "dashboards" or "visualizations" or both terms to describe a concept that more or less conveys the same meaning. The two terms are closely-related and therefore, the method assigns them to the same cluster. However, due to using them separately as well, their individual frequencies are somewhat lower and this probably results in not finding this cluster in Q1 (i.e., mainstream work), but in Q2 (i.e., peripheral work) in the initial analysis. In the merged dataset, the relative frequencies and co-occurrences of key-terms in the same papers have changed. This change is not due to importing bias to the dataset: during the merging of the lists, the key-terms that existed in both of them were replaced by only one, and additional terms were also considered for each paper. As such, the newly inserted terms resulted in changing the co-occurrences and the respective co-word frequencies, fetching more realistic results.

5.2 Phase 2: Maturation

The fact that a topic is identified as mainstream, under-developed, emerging or disappearing does not mean that the topic cannot change its state: it depends on what are the research interests at the given time, and it somewhat reflects the systemic dynamics at that particular moment. The maturation of the field is not illustrated on the strategic diagrams when they are read alone: it is also the detection of backbone topics that provides evidence on that aspect. The strategic diagrams explicate how the field is progressing (evolves), and on their own, they can only tell if there are motor themes leading the field. Accordingly, in contrast with some early studies that identified fragmentation of the field [8, 29], the findings of this study revealed backbone topics (Table 4) and motor themes (they belong to Q1 in Figure 7) and provide evidence for the gradual maturation of the field over the years. Those topics are the same that previous studies had identified as central (i.e., MOOCs, visualization, discussion forums, etc.) and now appear to be well-structured (dense), as well. Specifically, the core-periphery analysis (see Table 4) identified the most popular, core and backbone topics that emerged during the period 2011-2019. The backbone topics act as nodes that allow for the "information flow" through the whole network and facilitate the emergence of well-developed topics and their gradual transformation into core ones. Without those topics, the whole network would loose its cohesion and disintegrate into separate and unconnected topics. In the case of the learning analytics research field, the backbone key-terms (i.e., MOOCs, LMS, EDM, assessment, higher education, etc.) are also either motor themes (i.e., in Q1) and hold a central and dense position in the network), or they are bandwagon themes (basic and transversal, i.e., in Q4) and play a pillar role to the development of the field, as shown from the co-word analysis (Figure 7), signaling the years of maturation.

5.3 Phase 3: Challenges and implications

The rising question is, being data-informed about the mapping of the landscape of our community, where do we want to go next? Are we satisfied with the degree of completion of the conducted research on some areas or we need to allocate additional resources there? When we can safely claim that the research on a topic is mature enough so we can move to other emerging topics?

Eight (8) of the most popular themes are also core and backbone for the field, suggesting a high consistency between research interests and scientific efforts to maintain the sustainability of the field. As seen from those topics, learning analytics accumulated knowledge appears to be data-driven (e.g., MOOCs, ITS, EDM, visualization) and already grounded on a very specific research context. This finding is in line with and further confirms previous results [9] that the filed is lacking theoretical frameworks which can provide solid common grounds for further development. Failing to include theory and practice (e.g., pedagogical perspective, learning

theories) is likely to slow progress, fail to achieve cohesiveness and universality, and might threat validity [15]. Due to the nature of the learning analytics domain as a multidisciplinary research area at the intersection of data science, educational technology and learning science, in order to ensure cooperation among various disciplines and establish coherence among research themes with the aim to accelerate progress and maturation, the field needs new motor themes derived from already well-established knowledge (e.g., social learning, higher education, discourse analysis), that will have implication to the emerging research themes (e.g., classroom orchestration, eye-tracking, NLP, temporal analysis).

Furthermore, the issues raised in [37] concerning the adoption of learning analytics and in [9] regarding the limited implications for practice so far, need to be placed and seen from a realistic viewpoint: the problem with learning analytics adoption and utilization of its full potential is more likely to relate to various stakeholders and institutional readiness to engage in such endeavours rather than on the maturity of the research "*per se*" and the sophistication of methods. The limitation in this case may come from other disciplines (e.g., legal and ethical concerns). Indeed, as seen in Table 3, the community is working to that direction and topics related to privacy and policy are identified as peripheral, yet well-structured themes. Additional work is required, though, in order to contribute to making those topics central for the research community and to developing frameworks that could align the learning analytics scaling and adoption with the institutional policies.

5.4 Limitations

A limitation of the study is that the analysis includes only SoLAR publications (i.e., LAK and JLA), which despite being the flagship venues of learning analytics, as selection, bring some bias to the study (e.g., most of the papers are coming from USA, EU, Canada, and Australia [28]). To remove this selection bias, the analysis should also consider research published in other TEL conferences and journals. However, the results obtained from the content included in the analysis, present clear and valuable insights on learning analytics evolution seen through the lens of SoLAR publications.

Foremost, as already explained, the aim of this study was not to script the field or measure the impact of previous research: to claim impact, collecting and properly analyzing the effect sizes of previous work (i.e., meta-analysis) is within our future work plans.

REFERENCES

[1] R S Burt. 2004. Structural holes and good ideas. *American journal of sociology* 110, 2 (2004), 349–399.

[2] M Callon, J P Courtial, and F Laville. 1991. Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemsitry. *Scientometrics* 22, 1 (sep 1991), 155–205.

[3] M Callon, J-P Courtial, W A Turner, and S Bauin. 1983. From translations to problematic networks: An introduction to co-word analysis. *Information (International Social Science Council)* 22, 2 (1983), 191–235.

[4] M A Chatti, A L Dyckhoff, U Schroeder, and H Thüs. 2012. A reference model for learning analytics. *Intl J of Technology Enhanced Learning* 4, 5-6 (2012), 318–331.

[5] D Clow. 2013. MOOCs and the funnel of participation. In *3rd Intl Conference on Learning Analytics & Knowledge*. ACM, 185–189.

[6] M J Cobo, A G López-Herrera, E Herrera-Viedma, and F Herrera. 2011. Science mapping software tools: Review, analysis, and cooperative study among tools. *J. of the American Society for Information Science and Technology* 62, 7 (2011), 1382–1402.

[7] C Cuccurullo, M Aria, and F Sarto. 2016. Foundations and trends in performance management. A twenty-five years bibliometric analysis in business and public administration domains. *Scientometrics* 108, 2 (2016), 595–611.

[8] S Dawson, D Gašević, G Siemens, and S Joksimovic. 2014. Current state and future trends: A citation network analysis of the learning analytics field. In *4th Intl Conference on Learning Analytics & Knowledge*. ACM, 231–240.

[9] S Dawson, S Joksimovic, O Poquet, and G Siemens. 2019. Increasing the Impact of Learning Analytics. In *9th Conference on Learning Analytics & Knowledge*. ACM, 446–455.

[10] M de Laat, V Lally, L Lipponen, and R-J Simons. 2007. Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis. *J of Computer-Supported Collaborative Learning* 2, 1 (2007), 87–103.

[11] W Drachsler, H & Greller. 2016. Privacy and Analytics: It's a DELICATE Issue a Checklist for Trusted Learning Analytics. In *6th Intl Conference on Learning Analytics & Knowledge*. ACM, 89–98.

[12] E Duval. 2011. Attention please!: learning analytics for visualization and recommendation.. In *1st Conference on Learning Analytics & Knowledge*. ACM, 9–17.

[13] R Ferguson. 2012. Learning analytics: drivers, developments and challenges. *Intl J of Technology Enhanced Learning* 4, 5-6 (2012), 304–317.

[14] R Ferguson, T Hoel, M Scheffel, and H Drachsler. 2016. Guest editorial: Ethics and privacy in learning analytics. *Journal of learning analytics* 3, 1 (2016), 5–15.

[15] D Gašević, V Kovanović, and S Joksimović. 2017. Piecing the learning analytics puzzle: a consolidated model of a field of research and practice. *Learning: Research and Practice* 3, 1 (2017), 63–78.

[16] A Gifi. 1990. *Nonlinear multivariate analysis*. Wiley.

[17] W Glänzel. 2001. National characteristics in Intl. scientific co-authorship relations. *Scientometrics* 51, 1 (2001), 69–115.

[18] Q He. 1999. Knowledge Discovery Through Co-Word Analysis. *Library Trends* 48 (1999), 133–159. Issue 1.

[19] C-P Hu, J-M Hu, S-L Deng, and Y Liu. 2013. A co-word analysis of library and information science in China. *Scientometrics* 97, 2 (2013), 369–382.

[20] L Leydesdorff and L Vaughan. 2006. Co-occurrence matrices and their applications in information science: Extending ACA to the Web environment. *American Society for Information Science and technology* 57, 12 (2006), 1616–1628.

[21] G-Y Liu, J-M Hu, and H-L Wang. 2012. A co-word analysis of digital library field in China. *Scientometrics* 91, 1 (2012), 203–217.

[22] Y Liu, J Goncalves, et al. 2014. CHI 1994-2013: Mapping Two Decades of Intellectual Progress Through Co-word Analysis. In *32nd Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 3553–3562.

[23] R Mihalcea and P Tarau. 2004. Textrank: Bringing order into text. In *Conference on Empirical Methods in Natural Language Processing*. 404–411.

[24] B Misiejuk, K & Wasson. 2017. State of the Field on Learning Analytics. (2017).

[25] F Murtagh and P Legendre. 2014. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of classification* 31, 3 (2014), 274–295.

[26] M E Newman. 2005. A measure of betweenness centrality based on random walks. *Social networks* 27, 1 (2005), 39–54.

[27] A E Nielsen and C Thomsen. 2011. Sustainable development: the role of network communication. *Corporate Social Responsibility and Environmental Management* 18, 1 (2011), 1–10.

[28] X Ochoa and A Merceron. 2018. Quantitative and Qualitative Analysis of the Learning Analytics & Knowledge Conf. 2018. *Journal of Learning Analytics* 5, 3 (2018), 154–166.

[29] X Ochoa, D Suthers, K Verbert, and E Duval. 2014. Analysis and reflections on the third Learning Analytics & Knowledge Conference (LAK 2013). *Journal of Learning Analytics* 1, 2 (2014), 5–22.

[30] Z Papamitsiou and A A Economides. 2014. Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society* 17, 4 (2014).

[31] A Pritchard et al. 1969. Statistical bibliography or bibliometrics. *Journal of documentation* 25, 4 (1969), 348–349.

[32] P Rombach, M A Porter, J H Fowler, and P J Mucha. 2017. Core-periphery structure in networks (revisited). *SIAM Rev.* 59, 3 (2017), 619–646.

[33] T Schank and D Wagner. 2004. *Approximating clustering-coefficient and transitivity*. Universität Karlsruhe, Fakultät für Informatik.

[34] B A Schwendimann, M J Rodriguez-Triana, et al. 2017. Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies* 10, 1 (2017), 30–41.

[35] W-L Shiau, S-Y Chen, and Y-C Tsai. 2015. Management information systems issues: co-citation analysis of journal articles. *"Intl J of Electronic Commerce Studies"* 6, 1 (2015), 145–162.

[36] S Slade and P Prinsloo. 2013. Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist* 57, 10 (2013), 1510–1529.

[37] O Viberg, M Hatakka, O Bälter, and A Mavroudi. 2018. The current landscape of learning analytics in higher education. *Computers in Human Behaviour* (2018).

[38] H Waheed, S-Ul Hassan, N R Aljohani, and M Wasif. 2018. A bibliometric perspective of learning analytics research landscape. *Behaviour & Information Technology* 37, 10-11 (2018), 941–957.

[39] Z-Y Wang, G Li, C-Y Li, and A Li. 2012. Research on the semantic-based co-word analysis. *Scientometrics* 90, 3 (2012), 855–875.