Current State and Future Trends: A Citation Network Analysis of the Learning Analytics Field

Shane Dawson[1], Dragan Gašević[2], George Siemens[3], Srecko Joksimovic[4]

[1] Learning and Teaching Unit, University of South Australia, Australia
shane.dawson@unisa.edu.au

[2] School of Computing and Information Systems, Athabasca University, Canada
dgasevic@acm.org

[3] Learning Innovation and Networked Knowledge Research Lab, University of Texas, Arlington
gsiemens@uta.edu

[4] School of Interactive Arts and Technology, Simon Fraser University, Canada
sjoksimo@sfu.ca

ABSTRACT
This paper provides an evaluation of the current state of the field of learning analytics through analysis of articles and citations occurring in the LAK conferences and identified special issue journals. The emerging field of learning analytics is at the intersection of numerous academic disciplines, and therefore draws on a diversity of methodologies, theories and underpinning scientific assumptions. Through citation analysis and structured mapping we aimed to identify the emergence of trends and disciplinary hierarchies that are influencing the development of the field to date. The results suggest that there is some fragmentation in the major disciplines (computer science and education) regarding conference and journal representation. The analyses also indicate that the commonly cited papers are of a more conceptual nature than empirical research reflecting the need for authors to define the learning analytics space. An evaluation of the current state of learning analytics provides numerous benefits for the development of the field, such as a guide for under-represented areas of research and to identify the disciplines that may require more strategic and targeted support and funding opportunities.

Categories and Subject Descriptors
J.1 [Administrative Data Processing] Education; K.3.1 [Computer Uses in Education] Computer-assisted instruction (CAI)

General Terms
Measurement, Performance

Keywords
Citation analysis, author networks, learning analytics, social network analysis

1. INTRODUCTION
The increase in computational power and abundance of digital data has resulted in a growing interest in learning analytics and 'big-data' among educators, administrators and government bodies.

LAK '14, March 24 - 28 2014, Indianapolis, IN, USA
Copyright 2014 ACM 978-1-4503-2664-3/14/03 $15.00.

This interest stems from the potential for learning analytics to provide new insights and understanding into how students learn and how educators and institutions can best support this process. However, as noted in the McKinsey report, the strategic application of analytics [16] to inform practice has not been extensive within the education sector. Simply put, while the education sec- tor is data rich it has historically been analytics poor. The limited use of analytics in education, apart from the use of business intelligence to improve organizational efficiency, is clear in several areas. For instance, the lack of access to easy to use dashboard and analytics tools for educators and learners and the absence of any systemic deployment of learning analytics that informs the organization's learning and teaching practice. Arguably this dearth of institutional wide exemplars is in part due to the need for analysts and researchers to rapidly demonstrate the potential data mining and analytics to improve learning while dealing with the complexity of "messy" data and the organizational political challenges that arise due to various stakeholders and data "owners".

The outcome of limited systemic analytics activity is a predominance of research that is founded on the extraction of readily available data such as those drawn from learning management systems (LMS), student information systems (SIS), and basic demographics and student grades. Thus, the research questions have tended to centre on identifying key variables that inform student retention and academic performance. However, these commonly bivariate analyses are the "low hanging fruit" in terms of the overall potential for analytics to redefine and shape education praxis. The goal here is not to condemn prior analytics work to the sidelines of research but to note that the learning analytics landscape is complex, shrouded in the morass of social, technical, and cultural problems that pervade the education sector. While LMS and SIS data can provide insight into how to improve teaching and learning, this level of focus is not suitably aligned with the substantial challenges that face all levels of education – many of which require a systemic and integrated response. For example, while it is helpful to note that students who regularly log into a LMS may perform better than their less active peers, this information is not suitable for developing a focused response to poor performing students. It is neither helpful nor productive to simply tell under-performing students to log in more frequently. Despite the current limitations, much credit must be given to the early analytics work for rapidly advancing the profile of learning analytics and raising awareness about the value of educational data among a diverse set of stakeholders such as researchers, senior education administrators, government, industry, and funding bodies. LMS or SIS data can be a useful proxy for seeing a part of a problem, but it is insufficient to serve as a model for intervention that is based on the current state of learning sciences [1]. Learning analytics to date has served to identify a condition, but has not advanced to deal with the learning challenges in a more nuanced and integrated manner.

This paper is an evaluation of the current state of the field of learning analytics through the analysis of articles and citations. Literature in learning analytics under the broad umbrella of the Society for Learning Analytics Research (SoLAR) and Learning Analytics and Knowledge (LAK) conference now spans four years. A growing diversity of disciplinary approaches is evident in analytics literature, reflecting a field at the intersection of numerous academic disciplines. For instance, the field draws on assorted theory and methodologies from disciplines as diverse as education, psychology, philosophy, sociology, linguistics, learning sciences, statistics, machine learning/artificial intelligence and computer science. As the learning analytics field further evolves and matures it is important to reflect on the diversity of disciplines and methodologies that have contributed to the current state of learning analytics research. In so doing, there is a need to identify the emergence of trends and disciplinary hierarchies that may influence the future direction of the field.

An evaluation of the current state of learning analytics provides numerous benefits for the development of the field, including:

- a foundation for future research through the acknowledgement of past research activities;
- assistance for grant-making agencies by identifying promising research areas that align with regional and national education goals;
- identify disciplines that are under-represented and require more strategic and targeted support and funding opportunities;
- identify gaps in research for researchers and students; and
- improve the integration between theory and practice by identifying connections between researchers and papers.

This raises questions relating to the future impact and direction of learning analytics and how more granular and nuanced analytics activities can be deployed to facilitate uptake and application among an expanded set of stakeholders (e.g. research, government bodies, education administrators, technology support, and faculty). To address these questions this paper explores through the lens of structured mapping and citation networks the research domains and relationships that have extensively contributed to the field to date. An investigation of the patterns that evolve from citation networks can indicate the emergence of new inter- disciplinary research and methodological clusters alongside the identification of the more established and mature sub-communities. In undertaking a mapping and review of the collaborations that have evolved in the field, researchers and practitioners can identify the cliques and sub-culture that define the broader learning analytics community. As the 2014 LAK conference theme suggests, "*learning analytics at the cross roads,*" this is an ideal time to reflect and to assess if the field is moving to this intersection of research (and research disciplines) with learning theory and practice.

2. CITATION AND AUTHOR NETWORKS BRIEF OVERVIEW

The development and evolution of a scientific discipline can be considered as a complex system. As new knowledge is generated it is influenced by a wide and inter-related set of factors such as peer review processes, ethics, other closely aligned research disciplines and priorities in government funding to name but a few. Recently, the quantity and quality of the research activity within a discipline has been subject to a variety of measures to determine the overall impact and hence societal value. For example, as part of the research quality framework, the Australian government introduced the Excellence in Research for Australia as a process to assess the overall quality of university based research. Citation profiles and analyses are central measures adopted in these practices in order to determine research impact. This is founded on the basic notion that a core product of any research is the publication of scientific papers in journals and conferences. As the products of this research (papers) are developed they build upon and therefore cite other related and aligned research. Although there is much subjectivity regarding an individual's motivation to cite a particular author [6], pragmatically citation counts continue to be used as a proxy of impact. The core of the debate over the validity of citation analysis relates to the lack of assessment of the quality of the paper. As Waltman *et al*., [20] argued, highly cited papers are not always indicative of impactful research. However, as the authors further noted, on average this premise does tend to hold true. As such, it is reasonable to assume that high citation rates do reflect a certain level of quality [23].

Although there has been an increase in the use of citation and author analytics for quality assurance processes and benchmarking, these forms of analytics have long been adopted measures to indicate an author's or publication's relative influence and prominence within a research network [12]. In the late 1920's Gross and Gross [11] simply listed the citations made in the Journal of the American Chemical Society as a method to inform the strategic allocation of resources for journal acquisition in a library, specifically for the chemistry discipline. The conclusions drawn from this work can be seen to be an early process for determining the impact of an overall journal. In the 1950's Garfield's [8] ground-breaking work laid the foundation for contemporary work in citation analysis. Garfield established the "Science Citation Index" – a large and multi-disciplinary database now accessible via the "web of science".

Current citation databases such as Elsevier's Scopus, Google Scholar, web of science and others have made the extraction of co-citations and identification of co-author networks more accessible. As such, the development of citation networks have been adopted to identify disciplinary cliques or what Tight [19] refers to as the tribes and territories located within a broader research network. This data set is formed when for instance, Paper A refers to a previously published Paper B. This citation process can be seen to establish a tie between actors that ultimately forms a network.

There are multiple approaches to analyze this bibliometric data. For example, a network can be established through author to author ties or via direct citation. Co-authorship networks are generally adopted to provide an overview of the key contributors within a particular field. Collectively, authorship and citation analyses identify the distribution and accumulation of capital that develops as a research field evolves. Examination of these forms of capital can provide insight into the emergence of disciplinary hierarchies. Our aim for this paper is to provide a mapping of the learning analytics research community to identify the:

- prominent papers referred to in the research;
- dominant disciplines and methodologies adopted in learning analytics; and

- diversity of research paper genres that comprise learning analytics (e.g., opinion papers, reviews, conceptual, empirical research, etc.).

3. METHODS

This section details the data collection and analysis process undertaken in the study.

3.1 Data Collection

In this study, we included all papers that were published in the first three editions of the International Conference on Learning Analytics and Knowledge (LAK 2011, 2012, & 2013) [5; 9; 18]. The LAK conference was chosen as this is the major forum for the field of learning analytics and as such provides a solid foundation for examining the influential papers, authors, disciplines and methods that comprise learning analytics to date. To further compare and supplement this data set we examined three special issue journals. We identified three such special issues published since the first edition of the LAK conference – namely the Journal of Asynchronous Learning Networks (JALN 2012, vol. 16, no. 3, 2012); Journal of Educational Technology & Society (ETS 2012, vol. 15, no. 3, 2012) and the American Behavioral Scientist (ABS 2013, vol. 57, no. 10). In the following, we refer to these conferences and special issue journals as collected papers. A more comprehensive analysis comprising the entire collection of journal papers published on the topic is extremely complex and would require conducting a systematic review [13]. As the goal of our study was to provide an initial assessment of the emerging author and citation networks the publication sources was restricted to the conference and specific special issues occurring in the international journals. As such we also recognize that the extraction of authors and citations for the current study is an incomplete set of all research that has been undertaken within the broad scope of learning analytics. As these datasets become more complete we will be able to evaluate whether the findings presented here are an accurate reflection of the corpus or a result of the early analytic models employed and interpretations of this preliminary work.

From the collected papers, we extracted information on the authors of the research and the list of references cited per paper. Furthermore, we collected data relating to the home discipline for all listed authors of the collected papers. This was accomplished by examining the affiliation of the authors within the collected papers (e.g., if a researcher was affiliated with a computer science department, we assigned computer science as the home discipline for this author). In cases where such details were lacking in the collected papers we undertook searches through institutional websites, and social networking sites such as LinkedIn. Lastly, the methodology employed for each submitted paper was determined along with the papers genre (e.g., conceptual, empirical, or review papers).

3.2 Data analysis

The collected papers and the extracted information were analyzed using social network analysis and content analysis. Two social networks were established based on the collected information: i) author network and ii) citation network. The author network was created by including all authors as nodes and creating edges between co-authors. For example, if nodes A1, A2, and A3 co-authored paper A, we created nodes A1, A2, and A3 and established the following edges: A1 - A2, A1 - A3, and A2 - A3 (i.e., the author network was undirected). The citation network was formed by extracting the authors of both the collected papers and their citations. Edges were created between the authors of the collected papers and the authors of their citations. For example, if actors A1 and A2 co-authored a paper A, in which they cited paper B written by B1, B2, and B3, we created five nodes A1, A2, B1, B2, B3. Hence the following directed edges in the citation network were established: A1 - B1, A1 - B2, A1 - B3, A2 - B1, A2 - B2, A2 - B3. The nodes in the two networks were assigned the following attributes: i) home discipline, ii) type of research contribution of the co-authored papers, and iii) types of research methods used in the studies reported (see below for the details about the content analysis and the attributes under ii) and iii)). The three networks were analyzed using the following well-established measures in social network analysis [3; 4; 7; 22]:

- Degree – the number of edges a node has in a network
- Closeness – the distance of a node to all other nodes in the network;
- Betweenness – the number of shortest paths between any two nodes that pass via a given node;
- Eigenvector – the measure of influence of a node;
- Diameter – the maximum eccentricity of any node in a network;
- Modularity – a measure of decomposability of the network into modular communities.
- Clustering co-efficient – measure of completeness of the neighborhood of a node in a network when applied to a single node and is the average value of the clustering coefficients of all the nodes in the network.

All social networking variables were computed using the Gephi open source software for social network analysis [2]. As noted above content analysis was undertaken to determine two important research facets, namely: i) the types of contributions that the collected paper offered; and ii) the type of research methods adopted in the studies reported within the collected papers. For the analysis of the papers based on their contribution type, we followed the classification scheme commonly adopted in the areas of information systems and software technology [10; 17; 24]. The classification included the following categories: (1) Evaluation research provides an assessment of a particular problem and/or solution in practice and is typically based on research methods such as case study or field experiments; (2) Validation research investigates and proposes a novel technique or a solution that has not been not implemented in practice and is typically based on research methods such as rigorous analysis, experiments, and simulations; (3) Solution proposal is a novel or substantial extension of an existing method or technique and usually offers an example scenario of the proposal accompanied with a critical analysis of pros and cons of the proposal; (4) Conceptual proposal offers a new perspective to the phenomena under study and the structure of the field through a taxonomy or a conceptual framework; (5) Opinion-oriented papers offer positions on certain problems under study of broader interest for the field; and (6) Experience reports outline insights into the experience accumulated within projects. In addition, we extended this classification with the category of panel/workshop paper to be able to be inclusive of all categories within the LAK conference proceedings.

Analysis of the research methods adopted was undertaken using the following coding scheme: (1) qualitative method – indicating a well-established qualitative research method such as content analysis or grounded theory; (2) quantitative method – referring to well-established quantitative research methods such as hypothesis testing using statistical tests performed on the data collected through surveys, software logs, and social networks;

(3) mixed-methods – indicating a combination of qualitative and quantitative research methods; and (4) Other – includes papers such as reports of analyzed literature, or discussion of a particular topic or field more generally.

4. RESULTS AND DISCUSSION

4.1 Citation counts and network

Based on the extraction of the data detailed in the methods section above, the following counts of citations were calculated. The counts of individual citations are confined to a score of 1 regardless of how many times an individual paper may actually cite a specific publication. For example, reference may be made to a specific publication several times within a paper however this is weighted the same as an alternate publication with only a single mention. Examination of the number of citations a publication receives can reveal insight into the types and distribution of influential articles.

Table 1 lists the highest cited papers to date aggregated across the LAK conferences and special issue journals. Predictably given the stage of learning analytics research, the commonly cited papers are largely conceptual and review based. The exceptions being Macfadyen and Dawson's [14] empirical work and Wasserman and Faust's [21] oft cited book on SNA methodology. The reference to SNA methods reflects the volume of learning analytics re- search incorporating network analyses. The adoption of SNA techniques may also be indicative of researchers grounding their work in more socially oriented learning theory. However, further investigation is required to substantiate this claim. The high number of conceptual and review papers are indicative of the perceived need for authors to define and explain the learning analytics field in order to ground their work. As the field further evolves the need to re-establish a common understanding and definition of learning analytics will presumably decrease. In so doing, future analytics authors will increasingly draw upon more foundational and innovative empirical work. While direct citation counts are commonly used and provide a benchmark for future assessment, a richer more in-depth analysis is required. For instance, Ding *et al.*, [6] analyzed the number of and location within a publication where repeat citations occur. The authors argued that in order to determine influential papers in a field research should include the assessment of the number of re-cites and where these re-citations occur. Our future work in this area will seek to include these analyses to determine if repeat citations are significantly different in terms of identifying the key actors in the learning analytics landscape.

Table 2 provides an overview of the number and diversity of citations across the investigated publications. Calculation of the top 20 citations as a percentage of the total indicates this grouping only represents a small subset of the network. This finding suggests there is a high level of diversity in the references the authors are drawing upon. The 20-40 top citations band also accounts for a relatively low percentage across all the reviewed publication sites. This is to be anticipated given the relative immaturity of the LAK conferences and the field in general. As previously disparate disciplines are brought into the collective learning analytics space there is a re-orientation of discovery of the established literature. For example as authors with a strong computer science back- ground engage with the education literature there is a necessary phase of discovery within this domain. As such, the subsequent interpretation and application is therefore strongly influenced by this alternate home discipline. Other authors with a similar back- ground will presumably orient towards their more closely related network peers. This clustering effect can be determined in the network calculations.

Table 1: Top 10 most cited papers from LAK conferences and special issue journals (citations recorded as of October 2013)

# of LAK/ journal citations	Reference	Google Scholar Citations
16	Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an early warning system for educators: A proof of concept. Computers and Education, 54, 588–599.	90
15	Arnold, K. E. (2010). Applying academic analytics. Educause Quarterly, 33(1). Retrieved from http://www.educause.edu/ero/article/signals-applying-academic-analytics	57
14	Long, P. D., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education, EDUCAUSE Review, 46(5), 30-32, 34, 36, 38, 40. Retrieved from http://www.educause.edu/ero/article/penetrating-fog-analytics-learning-and-education	60
13	Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications. New York, NY: Cambridge University Press.	16680
11	Elias, T. (2011). Learning analytics: Definitions, processes and potential. Retrieved from http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf	37
11	Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. Expert Systems with Applications, 33(1), 135-146.	439
11	Dawson, S., Heathcote, L., & Poole, G. (2010) Harnessing ICT potential: The Adoption and Analysis of ICT Systems for Enhancing the Student Learning Experience. International Journal of Educational Management, 24, 2 2010, 116–128.	24
10	Baker, R. S. J. d., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. Journal of Educational Data Mining, 1(1), 3-17.	173
10	Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics: A new tool for a new era. EDUCAUSE Review, 42(4), 42-57.	120
10	Ferguson, R. (2012). The state of learning analytics in 2012: A review and future challenges (Tech. Rep. No. KMI-12-01). Milton Keynes, UK: Open University, Knowledge Media Institute. Retrieved from http://kmi.open.ac.uk/publications/techreport/kmi-12-01	53

Table 2: Citation overview

Publication	Total citations	Avg. / article	Total of Top 20 cited articles	Percentage of total citations (Top 20)	Total of Top 40 cited articles	Percentage of total citations (Top 40)
ABS13	455	52	52	11	89	19.6
ETS12	982	40.9	48	5	87	8.8
JALN12	164	23.4	30	18	50	30.5
LAK11	654	25.2	49	7.5	82	12.5
LAK12	946	19.7	62	6.5	102	10.8
LAK13	917	22	75	8.2	115	12.5
Average	686.3	30.5	52.7	9.4	87.5	15.8

Table 3: Citation network (based on a directed network)

	Nodes	Edges	Avg. degree	Average path length	Network diameter	Modularity	Avg clustering coefficient
LAK 11	1031	3109	3.0	4.02	8	0.75	0.47
LAK 12	1459	5001	3.43	3.58	9	0.81	0.5
LAK 13	1358	5684	4.19	3.10	9	0.76	0.48
LAK All	3133	13183	4.21	3.43	10	0.71	0.48
Journals	2281	7803	3.42	2.60	5	0.78	0.46

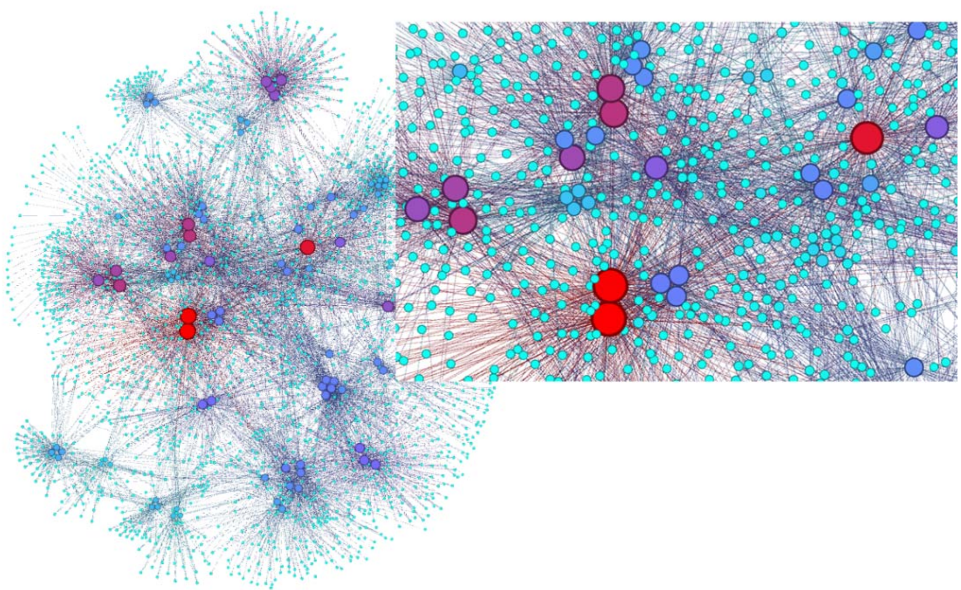

Figure 1a: All special issue journal citations. The network has been configured to highlight the clustering. Nodes sized by degree centrality

The clustering coefficient is a measure of the degree to which individual nodes in a network group (cluster) together (small worlds) [22]. A network clustering coefficient of 1 indicates that each node in the neighbourhood is completely connected. Conversely a score of 0 would indicate there are no connections. Analysis of all citations in the special issue journals revealed a moderate average clustering coefficient of 0.46 and a network diameter of 5 (Table 3). Calculation of the modularity score further indicates a high level of neighbourhood clustering.

The modularity score [3] in this instance was 0.78 at maximum resolution identifying 22 communities (Table 3). Interestingly when accounting for all citations made across the LAK conferences the clustering and modularity measures are very consistent with the special issue publications (Table 3).

However, the network diameter was observed to be greater with a score of 10. This suggests a more loosely connected network. Figure 1a and 1b illustrate the clustering occurring within the two networks, i.e. special issue journals and LAK conferences. Figures 1a and 1b also illustrate that the network appears to have few highly connected nodes. The degree centrality measure can be used to identify the highly cited nodes in the two networks. The degree centrality has been frequently used in bibliometric analyses to identify highly connected actors in a co-citation network. Identification of these nodes reveals researchers that tend to bridge both computer science domains and education. For instance the development of specific tools deployed and evaluated in the education space as well as educational psychology research using technologies.

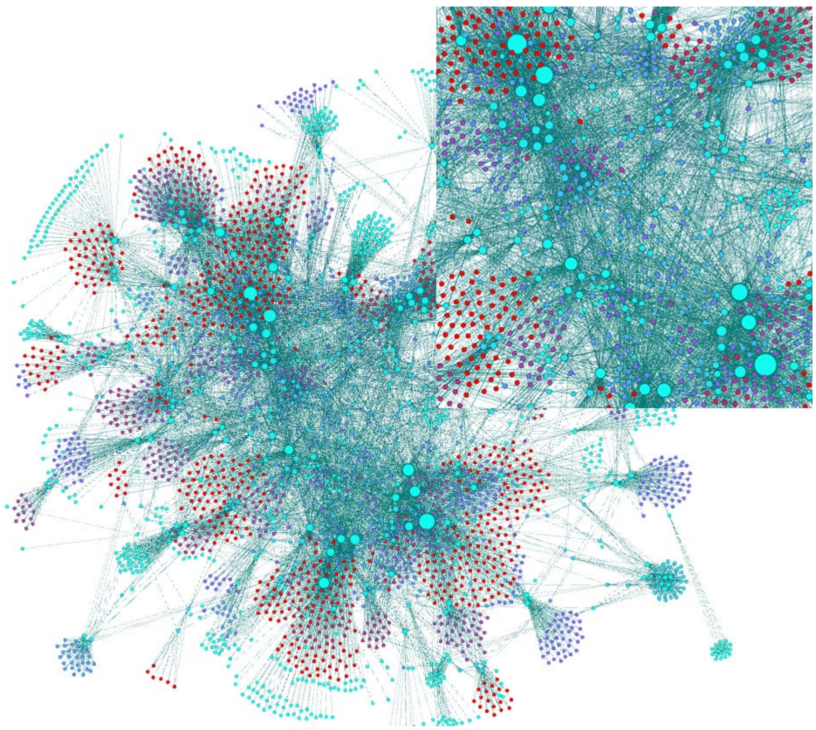

Figure 1b: All LAK conference citations. The network has been configured to highlight the clustering. Nodes sized by degree centrality

Table 4: Author network data

Author network	Nodes	Edges	Degree
LAK 11	55	67	2.436
LAK 12	126	183	2.905
LAK 13	141	234	3.319
LAK All	270	456	3.38
Journals	99	128	2.586

4.2 Author network

Authorship networks can be used to provide indicators of the diversity disciplinary backgrounds, and the thematic areas of interest associated within a field, in this case, learning analytics. For the purposes of this article we created a network, where a connection between two authors is established when they have a common paper in the investigated publication sites. The analysis of all authors represented across the LAK conferences demonstrates numerous small cliques with few highly interconnected authors. Figure 2 provides a representation of the author network for all LAK conferences where the majority of papers have 2-3 co-authors. This would appear smaller than the authorship profile in natural sciences for example where 3 or more are the norm [15]. While this paper presents a commencement point for these analyses in the learning analytics domain, future work should ideally draw in data from citation databases such as Scopus, and Google Scholar. This will provide a more complete assessment of the disciplinary connections that feed into learning analytics research. To this end, analysis of the disciplinary background of the authors can also be used to reveal the types of research that influences the field. Figure 3 illustrates the dominant disciplinary background of the authors represented in the LAK conferences. Clearly the conference is at present dominated by the computer science field representing approximately 51% of all authors. Education is also dominant with 40% of all authors having a background in this area. Several disciplines that would be expected to be prominent in learning analytics are limited their in influence to date including: machine learning, artificial intelligence, statistics, and data mining. Interestingly, at this stage of the learning analytics development there are relatively few inter-disciplinary nodes. However, as new researchers enter this space, and undertake their PhD and post-doctoral fellowships directly in learning analytics there will be a stronger identification of the juxtaposition of multiple disciplines and methodologies. Arguably, the researchers currently representing the field bring an established methodology and mindset influenced by their disciplinary theory, past experiences and approaches to research. The promotion of more inter-disciplinary teams and the emergence of new PhD candidates will result in new approaches and novel methods to tackle the key learning analytic challenges.

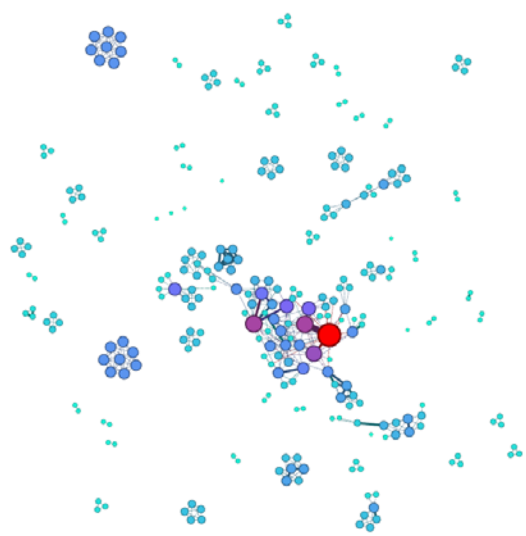

Figure 2: Network of all authors in the LAK conferences (nodes sized by degree centrality)

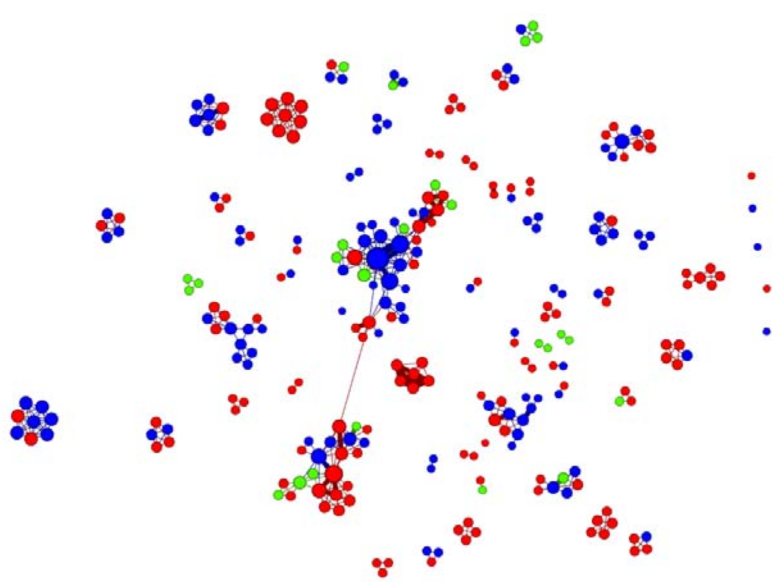

Figure 3: Network of all authors in the LAK conferences coded by disciplinary background. Red: Computer Science; Blue: Education; Green: Other (Industry, Engineering; Linguistics; or Business) (nodes sized by degree centrality).

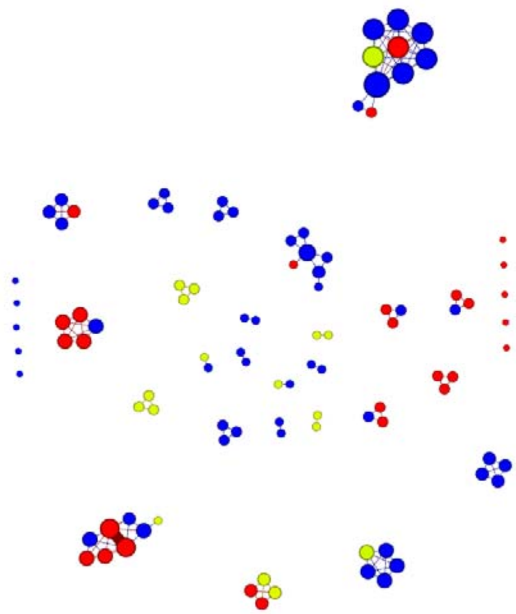

Figure 4: Disciplinary background of the authors represented in the special issue journals. Red: Computer Science; Blue: Education; Green: Other (Industry, Engineering; Linguistics; or Business) (nodes sized by degree centrality)

A similar comparison with the special issue journal suggests that while the computer science discipline dominates the LAK conference the journal publications are more popular for authors with an education research background. This finding may well reflect the differences in the publication impact viewed by the specific disciplines. For instance computer science researchers tend to place much value on the conference proceedings – in particular the partnership LAK has secured with the Association for Computing Machinery digital library (ACM). In contrast education researchers place a greater emphasis on journal publications - particularly a journal with an established and high impact factor. This separation is one of the organizational challenges that need to be addressed to avoid a fragmentation of the field. That is, how does Society for Learning Analytics Research (SOLAR), the major driver of the LAK conference, ensure the field does not separate into discipline dominated publication sites. Learning analytics is inter-disciplinary research and as such the conference proceedings and the newly established Journal for Learning Analytics research must provide an opportunity for researchers and practitioners to cross the boundary lines of their disciplines.

4.3 Practical implications

The analysis conducted in this paper can provide an understanding for how a field such as learning analytics evolves and matures by detailing the emergence of key authors, papers, thematics and the types of papers that are currently influencing the field. The intent of our analysis is not to offer a road map forward in the development of the discipline. Science and research can be influenced, but not scripted. As such, structured mapping and citation analyses serve primarily to raise awareness about the structure and attributes of knowledge in a discipline.

Numerous implications arise from our analysis, including:

- the development of curriculum in the growing number of academic programs that include learning analytics as a topic;
- promotion of under-represented groups and research methods to the learning analytics community;
- fostering the development of empirical work and decreased reliance on founding, overview and conceptual papers; and
- improved connections to sister organizations such as the International Educational Data Mining Society.

As indicated in Figure 5, the diversity of research contributions and the research method used by the authors across different disciplines is broad. It is clear that both the conference and the journal special issues are dominated by the lack of conventional research methods and that the authors, regardless of the home discipline, mainly contribute proposal solutions to the conference. They clearly indicate a much higher 'rigor' in the use of research methods by the researchers from education, who primarily use qualitative research methods. This is also corroborated by the higher numbers of papers in the category of evaluation research by the researchers from the education home discipline. Category other is the most common for research methods for the papers published in the journals. This category is now dominated by computer scientists and educators. This category is characterized by either reporting of descriptive statistics of (*ad hoc*) questionnaires, comparison of literature, or description of potential usage scenarios without actual data collection or formalization. The conference is dominated by proposal solutions authored by computer scientists and educators. The second most common category are evaluation research papers, which are dominantly authored by computer scientists followed by educators, who had almost an equal number of contributions in opinion papers, personal experience papers, as well as validation research. The second most important category for computer scientists was validation research. Again, the most common method for the conference is other – which is really not a formal method, but typically an ad hoc write up of personal experience, some random statistics and comparison with the literature. It is interesting that the conference has more dominated by the use of quantitative methods, which probably has to do with the higher number of computer scientist.

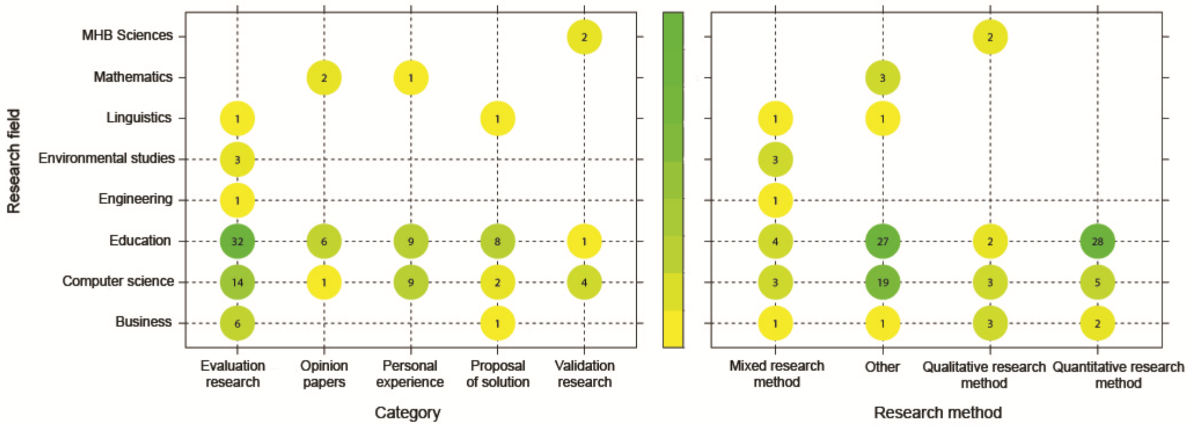

Figure 5a: Cross-tabulation of the authors' home disciplines (i.e., numbers in the circles represent the number of authors) with both research methods reported in the journal papers and types of journal papers.

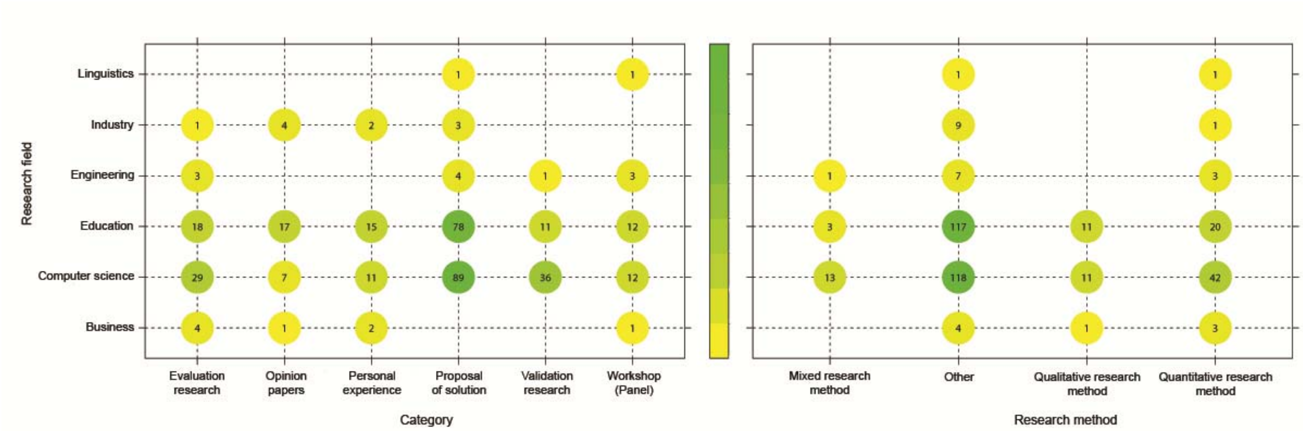

Figure 5b: Cross-tabulation of the authors' home disciplines (i.e., numbers in the circles represent the number of authors) with both research methods reported in the conference papers and types of conference papers.

5. REFERENCES

[1] Baker, R.S.J.D. and Siemens, G., 2013 In press. Educational data mining and learning analytics, Cambridge University.

[2] Bastian, M., Heymann, S., and Jacomy, M., 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. In Proceedings of the Third International AAAI Conference on Weblogs and Social Media (San Jose2009), AAAI Press.

[3] Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E., 2008. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008/10/P1008.

[4] Bonacich, P., 2007. Some unique properties of eigenvector centrality. Social Networks 29, 4, 555-564.

[5] Buckingham Shum, S., Gasevic, D., Ferguson, R., Haythornthwaite, C., and Dawson, S., 2012. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge ACM, New York, USA.

[6] Ding, Y., Liu, X., Guo, C., and Cronin, B., 2013. The distribution of references across texts: Some implications for citation analysis. Journal of Informetrics 7, 3, 583-592.

[7] Freeman, L.C., 1978. Centrality in social networks conceptual clarification. Social Networks 1, 3, 215-239.

[8] Garfield, E., 1955. Citation indexes for science: A new dimension in documentation through association of ideas. Science, New series 122, 108-111.

[9] Gasevic, D., Conole, G., Siemens, G., and Long, P.D., 2011. Proceedings of the 1st International Conference on Learning Analytics and Knowledge (LAK'11) ACM, New York, USA.

[10] Glass, R.L., Vessey, I., and Ramesh, V., 2002. Research in software engineering: an analysis of the literature. Information and Software Technology 44, 8, 491-506.

[11] Gross, P.L.K. and Gross, E.M., 1927. College libraries and chemical education. Science, New series 66, 1713, 385-389.

[12] Halevi, G. and Moed, H.F., 2013. The thematic and conceptual flow of disciplinary research: A citation context analysis of the journal of informetrics, 2007. Journal of the American Society for Information Science and Technology 64, 9, 1903-1913.

[13] Kitchenham, B., 2007. Guidelines for Performing Systematic Literature Reviews in Software Engineering. Keele University & University of Durham, Retrieved from http://cdn.elsevier.com/promis_misc/525444systematicreviewsguide.pdf.

[14] Macfadyen, L. and Dawson, S., 2010. Mining LMS data to develop an "early warning system" for educators: A proof of concept. Computers & Education 54, 2, 588-599.

[15] Mählck, P. and Persson, O., 2000. Socio-Bibliometric Mapping of Intra-Departmental Networks. Scientometrics 49, 1, 81-91.

[16] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A.H., 2011. Big Data: The Next Frontier for Innovation, Competition, and Productivity. McKinsey Global Institute.

[17] Petersen, K., Feldt, R., Mujtaba, S., and Mattsson, M., 2008. Systematic mapping studies in software engineering.

In Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering (Retrieved from http://www.bcs.org/upload/pdf/ewic_ea08_paper8.pdf2008).

[18] Suthers, D., Verbert, K., Duval, E., and Ochoa, X., 2013. Proceedings of the Third International Conference on Learning Analytics and Knowledge ACM, New York, USA.

[19] Tight, M., 2008. Higher education research as tribe, territory and/or community: a co-citation analysis. Higher Education 55, 5, 593-605.

[20] Waltman, L., Van Eck, N.J., and Wouters, P., 2013. Counting publications and citations: Is more always better? Journal of Informetrics 7, 3, 635-641.

[21] Wasserman, S. and Faust, K., 1994. Social Network Analysis: Methods and applications. Cambridge University Press, New York.

[22] Watts, D.J. and Strogatz, S.H., 1998. Collective dynamics of /`small-world/' networks. Nature 393, 6684 (06/04/print), 440-442.

[23] White, H.D., 1990. Author co-citation analysis: Overview and defense. In Scholarly communication and bibliometrics, C.L. BORGMAN Ed. Sage, Newbury Park, CA, 84-106.

[24] Wieringa, R., Maiden, N., Mead, N., and Rolland, C., 2006. Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. Requirements Engineering 11, 1, 102-107.