Author(s): Upreti, Bikesh & Asatiani, Aleksandre & Malo, Pekka

Title: To Reach the Clouds: Application of Topic Models to the Meta-review on Cloud Computing Literature

Year: 2016

Version: Post print

Please cite the original version:

Upreti, Bikesh & Asatiani, Aleksandre & Malo, Pekka &. 2016. To Reach the Clouds: Application of Topic Models to the Meta-review on Cloud Computing Literature. Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS). 10.

To Reach The Clouds: Application of Topic Models to the Meta-review on Cloud Computing Literature

Bikesh Raj Upreti
Aalto University
bikesh.upreti@aalto.fi

Aleksandre Asatiani
Aalto University
aleksandre.asatiani@aalto.fi

Pekka Malo
Aalto University
pekka.malo@aalto.fi

Abstract

Cloud computing remains an increasingly popular topic among practitioners as well as researchers. The literature spans across multiple disciplines, and the knowledge is fragmented and not systematized. To address this issue we apply topic models to conduct a meta-review on cloud computing. We identify twenty research topics across multiple disciplines, and demonstrate the use of the approach to conduct reviews in the field of information systems (IS). In additionally, we discuss multidisciplinary nature of cloud research, as well as research topics attracting contributions from various scientific fields.

1. Introduction

The topic of cloud computing is enjoying an increasing popularity among practitioners as well as researchers. According to the latest industry reports 69% of companies are planning to invest in the technology, and 80% of user companies report some degree of impact of cloud on their business [28]. At the same time number of published academic articles on cloud is growing exponentially [1].

Research on cloud computing is spread across multiple scientific disciplines including areas such as computer science, engineering, business administration and economics. While research focus may vary across disciplines, interlinks exist between different areas of cloud research. However, as cloud is a fairly new topic with an increasing body of literature, concept definitions and research efforts are fragmented. To address the problem of knowledge fragmentation reviews have been conducted concerning particular areas of cloud research (e.g. [1,10,16,30]). However, to the best of our knowledge no systematic effort has been made to conduct overarching review on cloud computing, which would include all research areas. Therefore, to address this gap, the main objective of our study is:

Objective 1: *What are specific research topics in cloud computing literature?*

The main challenge in conducting reviews on a topic, which encompasses multiple fields is a number of publications one would need to analyze. In order to accomplish our objective we apply a topic models method, which allows analyzing a large amount of textual data quantitatively. The topic models method has been successfully applied to such reviews in other fields [12,14,22], however, it has not been widely used in information systems [24]. Thus, secondary objective of this study is:

Objective 2: *Demonstrate application topic models to literature reviews on emerging areas in information systems.*

2. Background

It is widely believed that the use of the term "cloud computing" in the modern context began in 2006-2007 [25]. Since then cloud became an established concept in both industry and academia. There has been an exponential growth in cloud literature [1], which has been predicted by Yang and Tate [35] in an earlier review. However, there is a mismatch between foci of technology and business oriented studies [35].

Substantial work has been done to produce universal definition of cloud computing from the early days. Many definitions co-exist, each highlighting attributes of cloud relevant to a particular scientific discipline [18,20]. Table 1 presents examples from computer science, information systems and business literature, as cited in corresponding fields.

Table 1. Definitions of cloud computing from various perspectives

Reference	Definition	Field
Mell & Grance [21]	Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that	General

	can be rapidly provisioned and released with minimal management effort or service provider interaction.	
Youseff et al. [36]	A new computing paradigm that allows users to temporarily utilize computing infrastructure over the network, supplied as a service by the cloud provider at possibly one or more levels of abstraction	Computer science
Vouk [33]	Cloud computing embraces cyber-infrastructure, and builds on virtualisation, distributed computing, grid computing, utility computing, networking, and Web and software services.	Computer science
Foster et al. [11]	A large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted, virtualized, dynamically scalable, managed computing power, storage, platforms, and services are delivered on demand to external customers over the Internet.	Computer science
Buyya et al. [5]	A Cloud is a type of parallel and distributed system consisting of a collection of inter-connected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resource(s) based on service-level agreements established through negotiation between the service provider and consumers.	Computer science, information systems
Willcocks et al. [34]	The evolution of a service-based perspective on computing based on innovations in shared computing provision that improve simplicity, scalability, and efficiency.	Management

One of the most widely accepted definitions by National Institute of Standards and Technology defines cloud as a model of computing with defined set of characteristics [21]. More technically oriented literature outlines variety of infrastructural properties of cloud [5,11,36]. Whereas, business literature views cloud as a service paradigm abstracting it from any particular technological characteristics [34].

Such variance in understanding of the phenomena suggests two issues regarding cloud: 1) cloud is an emerging paradigm with growing body of academic literature; 2) there is disconnect across different disciplines resulting in fragmented research. Therefore we see an opportunity in conducting meta-analysis of previous scientific work across disciplines in order to categorize issues, and identify possible relationships between different research streams as well as gaps to study in the future.

A number of literature reviews exist on cloud computing covering various aspects of the research. These reviews can be divided roughly into four categories: 1) state of the cloud research in IS; 2) adoption and cloud sourcing; 3) cloud security and vulnerabilities; and 4) cloud migration and deployment.

Majority of literature reviews surveyed for the background of this paper addressed research in information systems field, with notable lack of reviews on studies published in technical fields. Earlier reviews discuss overall trends and definitions of the concept in IS and business studies [19,35]. In contrast, more recent studies highlight various aspects of the of cloud research, such as service composition [18], and client organization perspective [32].

Another group of reviews, published within business domain, focus on cloud sourcing and adoption. The reviews synthesize cloud adoption factors [1], compare sourcing determinants to traditional IT [30] and propose framework for the technology adoption [15]. Other reviews focus on particular uses of cloud such as healthcare [10] and government [31].

There is relatively limited number of reviews conducted on technology-oriented literature. From our survey we came across reviews on cloud security and vulnerabilities literature [13,16] as well as technical aspects on cloud migration [17].

3. Study

3.1. Data

To analyze topic structures in cloud computing literature, we analyzed bibliographic information, and article abstracts. We acquired data from Scopus, a bibliographic database by Elsevier. Scopus allows efficient collection of bibliographic information and abstracts in various file formats. Scopus also indexes majority of relevant journals in the fields of business, information systems and computer science. This makes Scopus an attractive choice as a data source for collecting data at a large scale.

Our data collection approach was based on step-wise elimination of irrelevant documents. At each step, we narrowed down document qualification criteria. We initiated search with “cloud computing” as a keyword in either abstract or in Title. The search result was then confined by publication date, from 2007 (a year cloud computing term was coined) to the first week of March 2015. The initial search yielded 19177 results, which was further narrowed down to only include documents in English language.

The subsequent search result reduced to 18571 documents and included sources like trade journal, book, trade publication and book series. Further, we opted for only peer reviewed scientific publications and thus filtered the search result by source type. The source filtering limited search result to 15079 documents. We also applied filtering by document type and excluded editorials, reviews, and notes to produce dataset of 13957 documents. The resultant data still covered a wide range of disciplines like biology, computer science, engineering, business, medicine and pharmaceutical. Thus, we introduced filtering based on subject areas to include business management and accounting, computer science, decision science, Economics, econometrics finance, engineering, mathematics and social science related publications only. The final search result contained 13396 documents with abstracts, article titles, authors, publication dates, source titles and citation counts. We further, removed 147 documents with incomplete information; 8, 16, 63, 24, and 36 documents with missing publication year, no author information missing, abstracts unavailable, incomplete abstracts and ambiguous abstracts, respectively. The final dataset comprised 13249 documents. Table 2 summarizes the process of data collection.

Table 2. Document Filtering Steps

	Operator	Filters	Results
1	No Operator	**Keyword:** "Cloud Computing" in "Article Title" OR "Abstract" AND Published Year > "2007"	19177
2	AND LIMIT-TO	**Language**: " English"	18571
3	AND LIMIT-TO	**Source Type:** "Conference Proceedings" OR "Journals"	15079
4	AND LIMIT-TO	**Document Type:** "Conference Paper" OR "Articles"	13957
5	AND LIMIT-TO	**Subject Area:** "Business Management and Accounting" OR "Computer Science" OR "Decision Sciences" OR "Economics, Econometrics, Finance" OR "Engineering" OR "Mathematics" OR "Social Sciences"	13396

3.2. Method

Researchers have been utilizing statistical methods such as principal component analysis, factor analysis, clustering algorithms and latent semantic indexing to identify topics being discussed in collections of literature. In these methods, a text document can be associated with only one topic or cluster at a time. However, document can exhibit multiple topics and heterogeneity in terms of main idea and subset of ideas that are being discussed. For instance, an article can discuss about hardware optimization algorithms for mobile application and its economic implications in cloud computing environment. Capturing such heterogeneity requires method that models a document as a multi-topic membership entity. In this study, we apply of a statistical method, Latent Dirichlet Allocation (LDA), as a model that captures intuition of multi topic characteristic of documents [2] in cloud computing literature.

LDA is the simplest topic model that is suited to for analyzing text documents [8]. Similar to all other latent variable models, LDA is based on assumption that complex observed data, i.e. a collection of documents, exhibits hidden, yet simple structures and patterns that are otherwise unnoticeable. The hidden patterns uncovered by LDA, referred to as topics, can be interpreted as a thematic structure that runs across the document collection [4]. The intuition of LDA is based on a mixed membership model that represents each document as a proportional mixture of topics. These topics are shared by all documents across the collection and can be viewed as a theme that runs throughout the collection and connects all the documents. Further, the proportional topic mixture provided by LDA can be interpreted as probability of topic memberships. The ability to model documents in terms of probabilistically multi-topic membership provides LDA with an edge over other classical mixture model [3].

3.3. Model

The first step in executing LDA is data preparation. Our input data combined both title and the abstracts from the cloud computing literature. Further, each document was given an identifier based on their year of publications. The next step involved selection of LDA tools. Among various available alternatives, we selected "Mallet package" in R Statistical environment (see http://cran.r-project.org/package=mallet). Mallet requires user to supply list of "stop words in English language" as a separate file. Stop words are common words used in English language that do not add any value in text analysis. Our list of stop words included single letters, numbers, articles, prepositions, pronouns, auxiliary verbs and their negations, and adverbs.

Implementation of LDA is based on the principle of hierarchical Bayesian inference. Before detailing the result, we briefly present mathematical formulation of LDA. We begin by laying out some

basic assumptions and notations. LDA assumes that a document is a "bags of words" without any orderings. It also assumes that documents in the collections are exchangeable i.e. order of documents can be ignored. A word, denoted by ***w***, is a basic unit of document and is assumed to be a discrete quantity. Similarly, a document is considered as a collection of ***N*** words arising from the vocabulary ***V***. Finally, a corpus is defined as a collection of ***L*** documents denoted by ***C***. LDA requires user to provide the number of topics, denoted by ***k,*** beforehand. To determine the number of topics, we relied on interpretability of topics to human readers. We tested our data with different numbers of topics ranging from 15 to 50 topics, with increment of 5 and evaluated the results. Based on our evaluation, we discovered that only 20 topics were meaningful and interpretable. Thus, we based our analysis on topic model with 20 topics as it fit our data.

The central computational problem of LDA involves inferring hidden topic structures given observed documents. Using hierarchical Bayesian approach, hidden topic structure can be computed from the interaction between the topic structure and documents given by their joint probability distribution. First, probability density of topic proportion based on given α prior is computed as:

$$p(\theta|\alpha) = \frac{\Gamma\left(\sum_{i=1}^{k} \alpha_i\right)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \; \theta_1^{\alpha_1 - 1} \dots\dots \; \theta_k^{\alpha_k - 1}$$

Where θ is proportion of topics for documents which represents probabilistic topic membership of all topics for given documents. Similarly, α is k dimensional positive vector containing scaling parameters, and Γ is a Gamma function. We used default α prior given by Mallet. Now using this information, the joint distribution of hidden topic variable and the observed data is sampled at three levels; word, document and corpus level. At word level, such joint probability distribution is computed for each word in all of the documents using following expression:

$$p(\theta, z, w|\alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^{N} p(z_n|\theta)\, p(w_n|z_n, \beta)$$ [1]

Where z is set topics for N words, i.e. topics variable from which w words are generated. In other words, z contains topic assignments for each of N words. Similarly, β is a V times k dimensional matrix of word probabilities for the topics, that is drawn from the Dirichlet distribution. It can also be interpreted as a matrix of most probable terms for each topics defined by corresponding word probabilities. To get the document level representation, the above quantity is first summed for all z values, i.e. for all words in document, and then integrated over θ, i.e. for all topics in the document. Mathematically:

$$p(w|\alpha, \beta) = \int p(\theta|\alpha) \left(\prod_{n=1}^{N} \sum_{z_n} p(z_n|\theta)\, p(w_n|z_n, \beta) \right) d\theta$$

The above expression gives the marginal probabilities for a single document. These values are computed for each document. The probability at corpus level is simply calculated by taking the product of marginal document probabilities.

$$p(C|\alpha, \beta) = \prod_{d=1}^{M} \int p(\theta_d|\alpha) \left(\prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn}|\theta)\, p(w_{dn}|z_{dn}, \beta) \right) d\theta_d$$

Based on the model described above, LDA utilizes posterior computational methods to compute unseen topic structures from the observed documents. The final outcomes of LDA include a topic word probability matrix β, a matrix with topic related terms and their corresponding probabilities for each topic, and document topic mixture matrix θ, a matrix of probabilistic topic assignments for all topics per document. For each topic, we generated 50 most probable terms of that topic. The other outcome per document topic matrix was normalized such that for a given document probability of belonging to all 20 topics summed up to 1.

The first task after achieving the results is giving meaningful name to the topics discovered. We began by generating topic labels using in-built function in Mallet that picks topic labels as per user specified number of terms with maximum topic belonging probabilities. We created topic labels using 3 most probable words for each topic. The topics discovered by topic models are more meaningful after human interpretation as human are more capable of assessing the cohesiveness and quality of topic discovered [9]. To add human interpretation to the topic labels, two authors separately named each topics based on the top 50 terms and abstracts of 10 articles with the highest topic probabilities. The final topic labels were based on the discussion that combined results from Mallet and author interpretations. The authors had unanimous agreement on the topic labels.

4. Results

Our sample consists of 13249 articles published from 2008 to March 2015. We also broke down publication statistics by year (see Figure 1). The number of publications per year has been increasing, peaking in 2013. Publications from 2015 include only the first three months. Therefore, it is impossible to conclude whether the decrease in annual publications

[1] http://cran.r-project.org/web/packages/mallet/mallet.pdf

is becoming a trend. Nevertheless, cloud computing remains a popular topic with a growing body of academic literature.

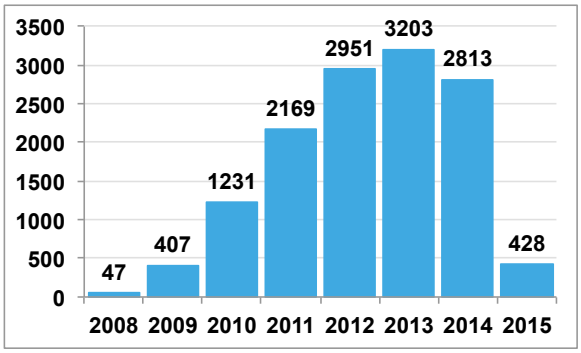

Figure 1. Publications on "cloud computing" per year in our sample

We also tracked publication outlets with the most articles published on cloud computing. Table 3 lists top ten outlets from our sample. Most popular outlets for publications remain cloud-specific conference proceedings. However, there are examples of journals with significant number of publications on cloud computing, such as Journal of Supercomputing and Journal of Theoretical and Applied Information Technology.

Table 3. Top 10 outlets by number of publications

Publication title	Articles
Future Generation Computer Systems	113
IEEE International Conference on Cloud Computing, CLOUD	89
Journal of Supercomputing	82
Proceedings - IEEE INFOCOM	72
Proceedings of SPIE - The International Society for Optical Engineering	61
Journal of Theoretical and Applied Information Technology	59
Procedia Computer Science	59
CLOSER 2011 - Proceedings of the 1st International Conference on Cloud Computing and Services Science	57
CEUR Workshop Proceedings	55
International Journal of Applied Engineering Research	55

In addition to the publication statistics, relying on citation data from the Scopus database, we also identified 10 most cited works from our sample (see Table 4).

Table 4. Most cited publications

Authors	Year	Citations
Buyya et al. [5]	2009	1483
Zhang et al. [37]	2010	526
Calheiros et al. [7]	2011	504
Nurmi et al. [23]	2009	432
Foster et al. [11]	2008	424
Satyanarayanan et al. [29]	2009	416
Ristenpart et al. [26]	2009	332
Marston et al. [20]	2011	324
Buyya et al. [6]	2008	316
Rochwerger et al. [27]	2009	255

4.1. Identified topics

The main result of this study is 20 topics identified by applying LDA to the sample of articles. Figure 2 present topic labels that combine independent evaluations from two authors and automated labels given by Mallet. The topics cover issues of deployment of cloud and cloud services in business; various aspects of privacy and security; issues related to cloud performance; and finally various practical applications of cloud.

Next we explored topics further by observing proportions of publications contributed by each topic. For each topic, we summed up proportion of a given topic in all documents and then normalized it by total number of publications. The final topic publication distribution is shown in Figure 2. The results show that the largest number of publications is related to *conceptualization of cloud*. The second d most popular topic is *cloud deployment models*. These two topics contributed more than 20% of the literature. Publications related to *cloud application development*, application of *cloud in manufacturing and transportations*, *cloud monitoring and testing and cloud optimization* are also among the major contributors.

We also ranked the topics in terms of impact on the cloud computing research. For these we used citations generated by each topic. To generate citation proportions for each topic, we used citation counts and probability loadings of each documents. Meaning that the total citation count of a document was distributed among 20 topics, based on the probability of the document to belong to a particular topic.

In these terms, *conceptualization of cloud* and *cloud deployment models* are on the top of the list. These works are frequently used as a fundamental reference material for majority of the papers on cloud computing. Topics such as *service provision and SLA, cloud monitoring and testing, data security and privacy*, and *cloud optimization* are also attributed above average citation proportions. The result of topic citation distribution is shown in the Figure 2.

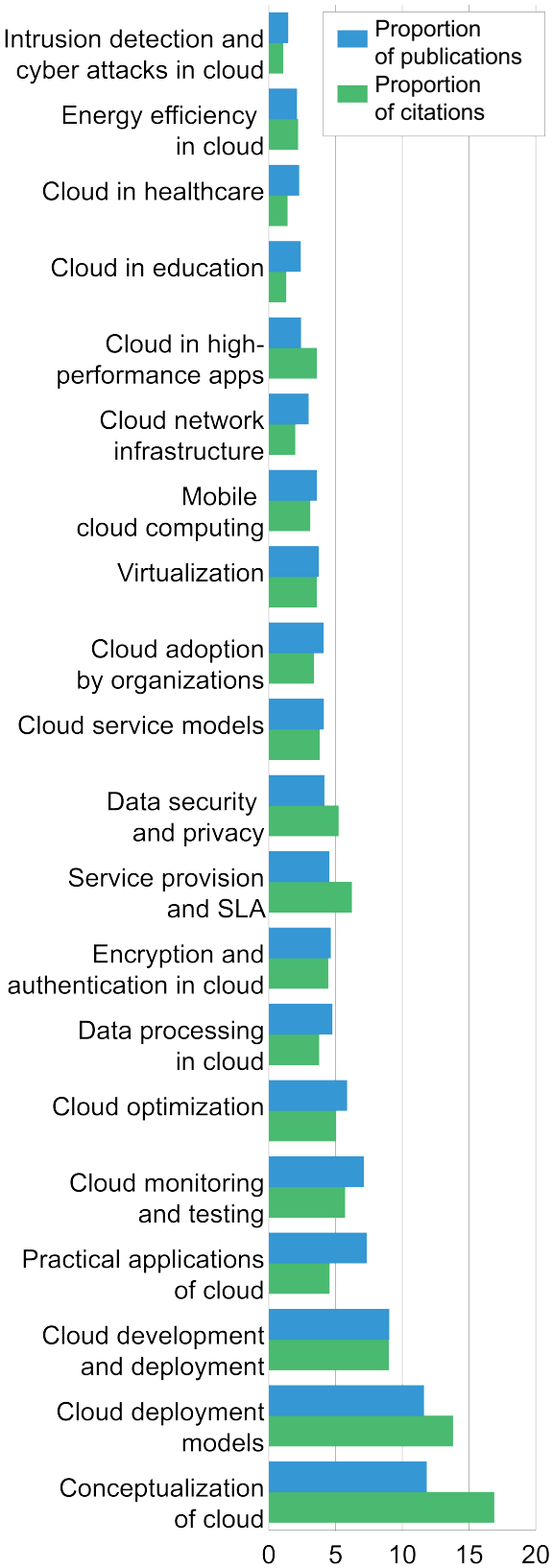

Figure 2. Topic labels, publication and citation proportions

4.2. Publication dynamics over the time

The proportion of publication does not provide the whole picture, as it lacks publication dynamics along the timeline. We also looked into how identified topics fared over the years in terms of number of publications. We aggregated topic probabilities per document for all topics on annual basis. The results depict the proportion of publications that a topic generated in a given year.

Figure 3 shows the proportion of publications on a time line scale for the major publication contributing topics. The results indicate that the proportion of publications on *conceptualization of cloud*, *cloud deployment models,* and *cloud applications development and deployment* is steadily decreasing. This can be partially explained by the fact that the fundamental definitions and base concepts are becoming universally accepted across various areas of research. Therefore, there is a decreasing demand for such work, and making significant contribution to the scientific discussion is difficult.

At the same time, Figure 3 demonstrates emerging topics, which have been steadily gaining popularity. Particularly, *cloud monitoring and testing,* and *cloud optimization* are worth mentioning. Emergence of these topics hints to the growing use of cloud computing, where performance of the system becomes an important issue.

Similarly, Figure 4 plots the time line for dynamic topics, which experience drastic changes in terms of proportion of publications. A number of publications related to *data security and privacy* is growing prominently. This may be related to the increasing attention towards data privacy following high-profile cases of government spying and data security breaches in major cloud providers. Topics of *cloud in high performance applications*, and *application of cloud in manufacturing and transportation* are experiencing a dip in the number of publications relative to other topics.

While interpreting topic time line development, we want to emphasize that these developments are relative to the overall growth cloud computing publications. The changes in publication proportions represent the annual share of publications for each topic. Therefore, it represents the interest of academic community in particular topic, relative to other topics, and not the absolute number of publications.

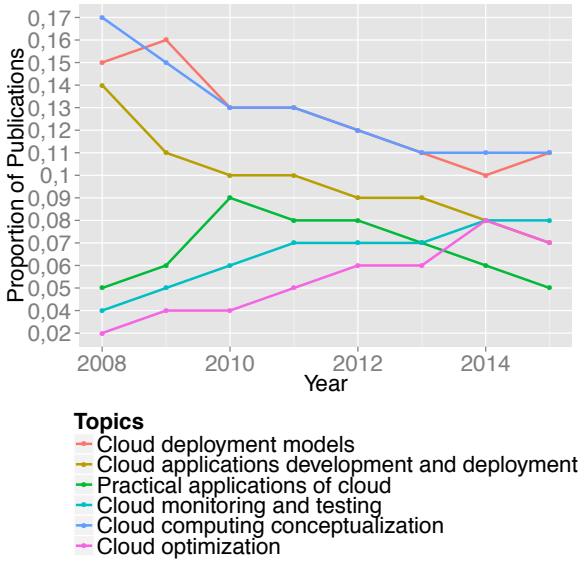

Figure 3. Publication timeline for six major topics

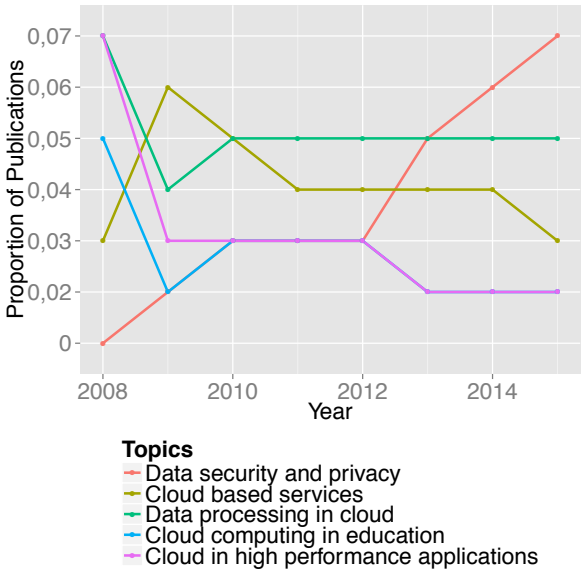

Figure 4. Publication timeline for five dynamic topics

5. Discussion

In the beginning of the paper we set two objectives for the study: 1) to identify specific research topics covered in cloud computing literature across multiple disciplines; 2) to demonstrate the application of topic models to literature reviews on emerging areas in information systems.

5.1. Multidisciplinarity in cloud research

We tackled the first objective by identifying 20 research topics in cloud computing literature, analyzing the contribution of each topic in terms of publications and generated citations, and presented insights concerning publication dynamics. Our results suggest maturation of the cloud computing as a research field. Scientific contributions are shifting from the descriptive and conceptualization work, towards studying applications of the technology, as well as peculiarities of its adoption and use.

In order to enrich the discussion, we explored multidisciplinarity of cloud research. To do this, we mapped all 20 topics across three areas of technical sciences, social science and information systems addressing socio-technical issues (see Figure 5). We manually analyzed the 20 titles, abstracts and publication outlets of articles with highest probability loadings for each topic.

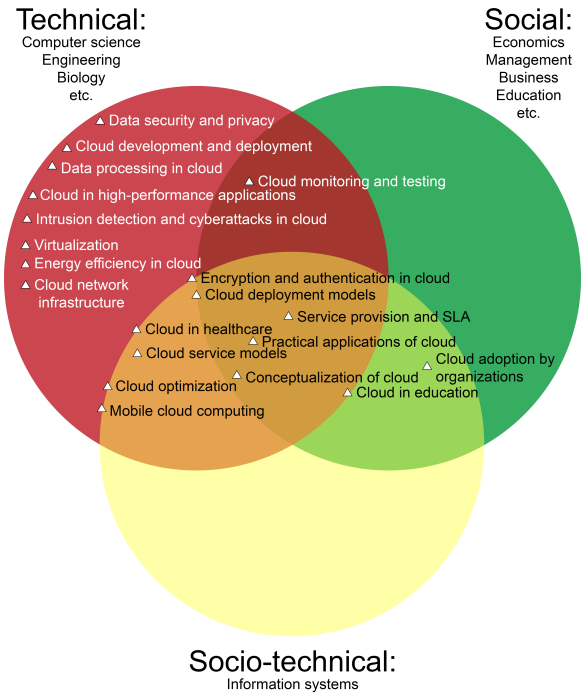

Figure 5. Illustration of topics organized by research areas

A big share of topics belongs purely to technical sciences, in particular to computer sciences. Some of the topics such as *energy efficiency* and *cloud in high-performance applications* include other technical disciplines, such as electrical engineering and mathematics. However, majority of the topics (12) are multidisciplinary, attracting contributions from both technical and social sciences.

Four topics attracted contributions from all three areas. Particularly notable are topics of *service provision and SLA,* and *practical applications of cloud. Service provision* attracts contributions from

scientific fields ranging from law and economics to human behavior and computer science. The topic of *practical applications of cloud* on the other hand covered various use cases and implications of cloud in a wide variety of contexts.

Results suggest that there is a potential for expanding disciplinary boundaries in topics situated between only two areas. A number of topics attracted contributions from only two areas. Notably *cloud in education* and *cloud adoption in organizations* are distanced from technical aspects of cloud, instead concentrating on social and socio-technical issues. On the other hand the topic *cloud in healthcare* is focused on computer science, biology and medical studies, with little proportion of issues such as usability or human-computer interaction in this context. Questions related to implications of cloud-based healthcare applications, on individuals behavior and society at large present opportunities for the future research.

The topics resulting from the study, dynamics of topic publication, and observations of multidisciplinary research suggest that cloud as a research field is shifting its focus from exploratory and conceptual inquiries encompassing wider range of scientific fields. Nevertheless, there remains a range of research opportunities in areas of social and socio-technical aspects of cloud technology.

5.2. Use of LDA for IS literature reviews

This study also demonstrates how topic models can be used for a meta-review, on emerging topics in IS, with a large number of published articles, thus accomplishing the second objective of our research. This method is by no means a substitute to narrative approaches of literature review, as it lacks analytical depth on specific publications. However, we believe that this approach could offer more depth into a research area, compared to the methods such as principal component analysis, when dealing with a large amount of bibliographical data. Further, unlike other methods, LDA can also be used for inferring topics of documents that were not part of the original dataset [3]. A new document in cloud computing can be probabilistically assigned to the existing 20 topics using existing terms distribution for the topics. The in-built inference functionality in Mallet provides an efficient way to incorporate new documents to existing collection. Thus, the utility of LDA can be extended to efficient management of literature.

5.3. Future research

In this study, we only scratched the surface on the big emerging area of cloud computing, identifying major research topics and presenting dynamics of the publications. Therefore, there are various directions which future research should concentrate on.

The next step from this study would be to conduct deeper analysis on multidisciplinarity of cloud research employing more robust quantitative techniques, to assign each article in the sample to specific research areas.

A number of topics are gaining momentum in cloud research, such as data security and privacy in cloud, or cloud optimization. To the best of our knowledge, there are no exhaustive reviews conducted on these topics, thus future research could focus on synthesizing the literature in these areas.

We demonstrated the use of topic models in a meta-review for IS, however we used particular approach of LDA to accomplish the task. The method can be further tested for different topics, and other approaches can be use, to improve validate the method.

5.4. Practical implications

While the study is primarily geared towards academic audience, there are implications for practitioners. Our analysis of publication dynamics reflects shifting focus of academics towards specific aspects of cloud. On the other hand, as disciplines of computer and information sciences strive to be on the cutting edge of technological development, shift in research interests, reflects the most important practical issues. Thus, this paper will help practitioners to track emerging trends in cloud research that are relevant to the industry. In addition, our research indicates that non-technical managerial skills are gaining importance for cloud deployment and use.

6. Conclusion

In this study we applied topic models to a meta-review of literature on cloud computing. We identified research topics across multiple disciplines, and demonstrated the use of the approach to conduct reviews in the field of IS.

As any research, this work has its limitations. First, the rigorous, quantitative categorization or articles, beyond identifying 20 topics was not included into the scope of this article. Second, we are relying on the Scopus database as a source of information, which has inherent shortcomings, such as cases of incomplete bibliographical and citation information. Further, LDA requires users to determine number of topics, which is a potential source of bias. And finally, approach used in this review is based on statistics. The topic models method allows analyzing a large amount of data, and systematic mapping of literature. However, the

method also lacks depth of narrative reviews, which allow addressing a broader range of research questions.

7. References

[1] Asatiani, A.Why Cloud? - A Review of Cloud Adoption Determinants in Organizations. *Proceedings of the 23rd European Conference on Information Systems (ECIS)*, (2015), 1–17.

[2] Blei, D.M. and Lafferty, J.D.Topic models. *Text mining: classification, clustering, and applications 10*, (2009), 71.

[3] Blei, D.M., Ng, A.Y., and Jordan, M.I.Latent Dirichlet Allocation. *Journal of Machine Learning Research 3*, (2003), 993–1022.

[4] Blei, D.M.Introduction to Probabilistic Topic Modeling. *Communications of the ACM 55*, (2012), 77–84.

[5] Buyya, R., Yeo, C.S., Venugopal, S., Broberg, J., and Brandic, I.Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. *Future Generation Computer Systems 25*, 6 (2009), 599–616.

[6] Buyya, R., Yeo, C.S., and Venugopal, S.Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities. *High Performance Computing and Communications, 2008. HPCC'08. 10th IEEE International Conference on*, (2008), 5–13.

[7] Calheiros, R.N., Ranjan, R., Beloglazov, A., De Rose, C.A.F., and Buyya, R.CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms. *Software: Practice and Experience 41*, 1 (2011), 23–50.

[8] Chaney, A.J.-B. and Blei, D.M.Visualizing Topic Models. *ICWSM*, (2012).

[9] Chang, J., Gerrish, S., Wang, C., and Blei, D.M.Reading Tea Leaves: How Humans Interpret Topic Models. *Advances in Neural Information Processing Systems 22*, (2009), 288–296.

[10] Ermakova, T., Huenges, J., Erek, K., and Zarnekow, R.Cloud Computing in Healthcare–a Literature Review on Current State of Research. *AMCIS 2013 Proceedings*, (2013).

[11] Foster, I., Zhao, Y., Raicu, I., and Lu, S.Cloud Computing and Grid Computing 360-Degree Compared. *2008 Grid Computing Environments Workshop*, (2008), 1–10.

[12] Hall, D., Jurafsky, D., and Manning, C.D.Studying the History of Ideas Using Topic Models. *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, (2008), 363–371.

[13] Hashizume, K., Rosado, D., Fernández-Medina, E., and Fernandez, E.An analysis of security issues for cloud computing. *Journal of Internet Services and Applications 4*, 5 (2013), 1–13.

[14] He, Q., Chen, B., and Giles, C.L.Detecting Topic Evolution in Scientific Literature : How Can Citations Help ? *Cikm*, (2009), 957–966.

[15] Hoberg, P., Wollersheim, J., and Krcmar, H.The Business Perspective on Cloud Computing - A Literature Review of Research on Cloud Computing. *AMCIS 2012 Proceedings*, (2012).

[16] Iankoulova, I. and Daneva, M.Cloud Computing Security Requirements : a Systematic Review. *Sixth International Conference on Research Challenges in Information Science (RCIS)*, IEEE (2012).

[17] Jamshidi, P., Ahmad, A., and Pahl, C.Cloud Migration Research: A Systematic Review. *IEEE Transactions on Cloud Computing 1*, 2 (2013), 142–157.

[18] Jula, A., Sundararajan, E., and Othman, Z.Cloud computing service composition: A systematic literature review. *Expert Systems with Applications 41*, 8 (2014), 3809–3824.

[19] Madhavaiah, C., Bashir, I., and Shafi, S.I.Defining Cloud Computing in Business Perspective: A Review of Research. *Vision: The Journal of Business Perspective 16*, 3 (2012), 163–173.

[20] Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., and Ghalsasi, A.Cloud computing — The business

perspective. *Decision Support Systems 51*, 1 (2011), 176–189.

[21] Mell, P. and Grance, T.The NIST Definition of Cloud Computing, Recommendations of the National Institute of Standards and Technology. *National Institute of Standards and Technology*, (2011).

[22] Newman, D., Noh, Y., Talley, E., Karimi, S., and Baldwin, T.Evaluating Topic Models for Digital Libraries Categories and Subject Descriptors. *Jcdl*, (2010), 215–224.

[23] Nurmi, D., Wolski, R., Grzegorczyk, C., et al.The eucalyptus open-source cloud-computing system. *Cluster Computing and the Grid, 2009. CCGRID'09. 9th IEEE/ACM International Symposium on*, (2009), 124–131.

[24] Okoli, C. and Schabram, K.*A Guide to Conducting a Systematic Literature Review of Information Systems Research*. 2010.

[25] Ragalado, A.Who Coined 'Cloud Computing'? *Technology Review*, 2011. http://www.technologyreview.com/news/425970/who-coined-cloud-computing/.

[26] Ristenpart, T., Tromer, E., Shacham, H., and Savage, S.Hey, you, get off of my cloud: exploring information leakage in third-party compute clouds. *Proceedings of the 16th ACM conference on Computer and communications security*, (2009), 199–212.

[27] Rochwerger, B., Breitgand, D., Levy, E., et al.The reservoir model and architecture for open federated cloud computing. *IBM Journal of Research and Development 53*, 4 (2009), 1–4.

[28] SAP and Oxford Economics.*The Cloud Grows Up*. 2015.

[29] Satyanarayanan, M., Bahl, P., Caceres, R., and Davies, N.The case for vm-based cloudlets in mobile computing. *Pervasive Computing, IEEE 8*, 4 (2009), 14–23.

[30] Schneider, S. and Sunyaev, A.Determinant factors of cloud-sourcing decisions: reflecting on the IT outsourcing literature in the era of cloud computing. *Journal of Information Technology*, (2014).

[31] Tsaravas, C. and Themistocleous, M.Cloud Computing and eGovernment: A Literature Review. *European, Mediterranean & Middle Eastern Conference on Information Systems*, (2011), 154–164.

[32] Venters, W. and Whitley, E.A Critical Review of Cloud Computing: Researching Desires and Realities. *Journal of Information Technology 27*, 3 (2012), 179–197.

[33] Vouk, M.Cloud computing — Issues, research and implementations. *ITI 2008 - 30th International Conference on Information Technology Interfaces*, (2008), 31–40.

[34] Willcocks, L., Venters, W., and Whitley, E.A.*Moving to the Cloud Corporation: How to Face the Challenges and Harness the Potential of Cloud Computing*. Palgrave Macmillan, 2013.

[35] Yang, H. and Tate, M.A descriptive literature review and classification of cloud computing research. *Communications of the Association for Information Systems 31*, 2 (2012), 35–60.

[36] Youseff, L., Butrico, M., and Silva, D. Da.Toward a Unified Ontology of Cloud Computing. *Grid Computing Environments Workshop*, (2008), 1–10.

[37] Zhang, Q., Cheng, L., and Boutaba, R.Cloud computing: state-of-the-art and research challenges. *Journal of Internet Services and Applications 1*, 1 (2010), 7–18.