Distribution Agreement

Signature:

______________________________ ______________

Name: Zhu Zeng Date 04/10/2019

Trends in Consumer Research Literature: A Historical Analysis

By
Zhu Zeng
Master of Business Studies
Marketing

____ [Advisor's signature]
Douglas Bowman, Ph.D.
Advisor

____ [Member's signature]
Ryan Hamilton, Ph.D.
Committee Member

____ [Member's signature]
Jagdish N. Sheth, Ph.D.
Committee Member

Accepted:

____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

____
Date

Trends in Consumer Research Literature: A Historical Analysis

By
Zhu Zeng
B.S., Emory University, 2013

Advisor: Douglas Bowman, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Masters of Business Studies
2019

Abstract

Trends in the Consumer Research Literature: A Historical Analysis

Authors' choices of words and subject terms arise from generative processes that include the underlying structures. These hidden topics and the distributions over words could reflect the thoughts and factors that drive innovation and changes in the field. The goal of this research is to uncover the hidden thematic structure that lives inside the collection of observed words and to investigate the evolution of substantive topics over the years. The data used in this analysis comprise articles published in the Journal of Consumer Research (JCR) and Advances in Consumer Research Proceedings (ACR) (1974–2017). There are 14,286 articles with more than 10 million words. A dynamic topic model captures the topic evolution. The way authors talked about a topic in 1974–1984 differs from the same in 2007–2017. Words underlying topics have probability changes over time. The interaction between the two publications JCR and ACR tends to have the smallest distance when they are close to each other in time, whereas a single topic might fail to have a powerful influence on the same topic across these two publications.

Trends in Consumer Research Literature: A Historical Analysis

By
Zhu Zeng
B.S., Emory University, 2013

Advisor: Douglas Bowman, Ph.D.

A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Business Studies
in Marketing
2019

Acknowledgement

I would like to express my heartfelt gratitude to my advisors, Dr. Bowman for his guidance. I couldn't have gone through the difficult time without his help and support in summer 2018. Thanks for being my advisor and guiding me all the way. I would also like to sincerely thank my committee members, Dr. Sheth and Dr. Hamilton for their advice in helping me with my work.

Table of Contents

1, INTRODUCTION

The words and subject terms that authors choose arise from certain generative processes, which include underlying structures. These hidden topics and their distribution over words could reflect the thoughts and factors that drive innovation and changes in the field. Based on this view, we are looking to uncover the hidden thematic structure within the collection of observed words and to investigate the evolution of substantive topics over the years. We also wish to study how the development of topics in Advances in Consumer Research (ACR) Proceedings might relate to the development in topics at Journal of Consumer Research (JCR). Having a shorter review process and being often an outlet for new ideas make ACR Proceedings a possible leading indicator of more robust studies later published in JCR.

There is a long history of investigations into how a field has evolved in academic journals in substantive disciplines, including that of consumer research. Nocosia (1963) examined the research outline in the history of consumer research from 1900 to 1950. Helgesson's (1984) content analysis focused on the growth of consumer behavior research from 1950 to 1981 and classified several key variables from selected journals and proceedings. Trends in consumer literature were examined by Kollat, Blackwell, and Engel (1972). Many studies from different disciplines are survey-based or citation-based (Leeflang et al. 2000, Lilien 1994, Buzzell 1968, Lilien et al. 1992, Baumgartner and Pieters 2003, Tellis et al. 1999, Moussa and Touzani 2010, Hall CM 2006, McKercher 2008). In the area of marketing, citation analyses have also been used by Baumgartner (2010) to address questions relating to the history of consumer research. Some studies focus on mechanisms, outcomes, and relationships pertaining to the authors' reviews, journal ratings, and rankings (Howey et al. 1999, Jamalet al. 2008, McKercher 2005, McKercher et al. 2006, Hall 2011, Ryan 2005).

The current study is more empirically grounded. Although we cannot compute the hidden structure, we can approximate it by modeling the collection of abstracts probabilistically. There is a class of analysis tools (topic models) in the computer science field. Several approximation techniques have been developed for Latent Dirichlet Allocation (LDA) in recent years: mean-field variational methods, expectation propagation, collapsed Gibbs sampling and collapsed variational inference (Blei et al. 2001, 2003, Minka and Lafferty 2002, Griffiths and Steyvers 2002, Teh et al. 2006). Among these techniques, collapsed variational inference[1] uses the best of mean field variational methods, and collapsed Gibbs sampling is a very efficient inference technique (Blei et al. 2009). To date, LDA has been applied to a number of academic journals in computer and social sciences (as shown in Table 1).

Most studies in social science use traditional LDA. For example, Wang, Bendle, Mai, & Cotte (2015) reviewed 40 years of JCR. Kevork and Vrechopoulos (2009) considered the keywords used in customer relationship management research. Mela, Roos, & Deng (2013) investigated the history of keywords used in marketing science.

Although traditional LDA can catch the evolution of topics by calculating their probabilities over different periods within certain documents, it assumes that the documents are exchangeable. The distribution or composition of words under the same topic in different times is assumed to be static. The problem with this is that the same topic may change over time. For example, the topic of "advertising" a few decades ago may have contained words like "affect," "impact," and "risk" with high probabilities. Nowadays, the same topic may include words like

[1] Collapsed variational inference was used in current study

"custom" and "social" with growing probabilities. Hence, although the topic remains unchanged, the words underlying that idea may change over time in ranking and probabilities. Besides, new words could be added and others could fade away. Similarly, the way a topic is addressed in earlier articles could differ from how it is addressed in later articles. Therefore, documents in time sequence are highly likely not to be exchangeable.

Unlike traditional LDA, dynamic topic models (DTM) can not only calculate an extensive document collection efficiently but also allow topic distributions to evolve from time slice to time slice (Blei 2016). DTM allows for a more reasonable study of the development of topics. We used this model in the current study and further investigated the evolution by adding an interaction of JCR with conference papers. Several findings emerge.

First, both JCR and ACR top words in four time slices align with common practical frameworks. Elements that relate to the marketing mix "price," "advertis" (promote), "product,” remain popular over time. However, our findings suggest that not all top words have always been popular. Some of them are in growth and some die away. The use of "social" has increased in both ACR and JCR in the last decade. The same changes in JCR as well as ACR may be associated with environmental changes (societal, technological, economic, public policy, etc.) in this era.

Second, the DTM model captured something salient about the evolution of topics in abstracts published through time. We noticed that the probability of words underlying topics has changed over time: that is, the way a topic was talked about in 1974–84 differs from how a topic was talked about in 2007–2017. Third, although similar to Wang et al. (2015), we found the topics of advertising, buying process, self-control, family decision making, memory price and price associations, resource constraints, social identity and influence, and methodological issues, we found some different topic distributions (topic weight) over time.

As for the interaction between JCR and ACR, we uncovered that they tend to have the smallest distance when they are close to each other in time. However, a single topic might fail to have a powerful influence on that same from ACR to JCR in the following year.

The rest of the article is organized as follows: In the second section, we detail the data structure used in the current study. Next, we outline our analysis approach and theory. Following that, we present our observations and findings. We then conclude this article by summarizing our key insights, limitations, and possible future research directions.

2, DATA

2.1 Articles

We scraped all articles published in ACR from its inception in 1974 to the most recent volume (Volume 45, year 2017) from the ACR proceedings website (with ACR's permission), using BeautifulSoup (acrwebsite.org/web/conferences/proceedings.aspx). Among these 45 volumes, the documents for 15 volumes (V31~V45) are stored only as pdf versions on ACR, so we converted these files into text versions (using PyPDF2 and pdfminer). A few pdf links contained blank information, so we deleted them from our data collection. We were left with 12,276 articles. We only used the main articles in the pdf and removed references, job titles and positions. Moreover, we deleted common phrases or words in ACR, such as "advances in consumer research volume," "http www acrwebsite org," "this work is copyrighted by the association for consumer research," etc. Some words have got stuck together after the pdf to text conversion, but these account for a small proportion of the entire collection. We corrected some of these and deleted the remaining ones when we were removing rare words.

We downloaded all articles published in JCR from its inception in 1974 through 2018 from Web of Science and we collected Journal of Consumer Psychology (JCP) from 2000 to 2017 (leaving out 1992–1999). Since ACR (North American Advance) has 45 volumes on their website so far (1974–2017), which we used as our ACR data input, we kept JCR only from its inception in 1974 through 2017 (44 volumes, Volume 1~Volume 44, Issue 4) and saved JCP for future studies. There are 2010 JCR articles remained after removing editors' specials and articles without abstracts from these 44 volumes.

To account for the power-law of word usage, we removed punctuation, numbers, and stopwords ("and," "the," etc.). We used NLTK's Stopwords and MySQL's Stopwords to clean the data. As Wang et al. (2015) had also excluded the words that were widely used in most topics, we also cleaned our data by removing words that were frequently used in most consumer research topics. Such words included "result," "study," etc.; a word list can be found in the code in our appendix). Since the total number of articles in our data set is 14,286 (JCR and ACR combined), there are more than 10 million words. Wang et al. (2015) did not aggregate words, like Mela et al. (2013) did, or change formation (e.g., buyer, buy, buying, purchase, shopping, child, children, etc.). We also did not aggregate our 10 million words. Rare words (with a frequency < 10) were removed to speed up the computation and a total of 1,455 unique words remained.

We first took all the articles together and fitted the models with random distributions of topics over words that generated those articles. Then we split them into slices to examine the evolution using DTM. ACR has one volume each year, except for the year 2011, which has two volumes (V38 and V39). JCR has 44 volumes in total and both ACR and JCR span from 1974 to 2017. Therefore, we divided this corpus into four sequential slices by year in both JCR and ACR (2×4), with nearly one decade per slice. Each slice contains 11 years (1974–1984, 1985–1995, 1996–2006, and 2007–2017). Since ACR has two volumes in 2011, the last slice of ACR contains 12 volumes. We assumed that documents within each slice are exchangeable but that documents outside the slice are not (e.g., 1974 and 1975 are exchangeable but 1974 and 1988 are not).

2.2 Stemmer

Schofield A & Mimno D (2016) trained and evaluated topic models using several different stemming algorithms. One of their findings was that stemmers do not produce meaningful improvement in likelihood and coherence. They found that stemmers could actually degrade topic stability. Stemming could sometimes result in fewer possible models, since multiple words with different or multiple meanings could be reduced to one token. A robustness check was conducted by fitting the model without stemming the words, and we found that they resulted in similar topics. Stemming the words after modeling saves space for different tokens and allows reviewers to browse the entire list easily. In addition, post-stemming is computationally cheaper. Hence, we post-stemmed the list of words.

2.3 Number of Topics

We treated our data set as observations that arise from a generative process that includes things we cannot observe. The only thing that we observe is the words of each document in our collection: this is all the information we have. Hence, we did not know the number of topics and we did not choose the number of topics at first. We used five-fold cross-validation to decide the number of topics. This method divided the data into five different numbers of subsets. Each data could get one chance to be in the validation. It took turns to use one subset as the validation set and the other four sets as training sets. Given the total of 14,286 documents, there could be many candidate numbers of topics. Hence, we did the cross-validation with different values of the number of latent topics in order to estimate the ideal number of topics. We first started with (2, 3, 4, 5, 10, 16, 20, 30, 40, 50, 75, 100) and then used (2, 3, 4, 5, 10, 16, 20, 30) as the candidate number set. Perplexity is a measure of how well a probability model predicts a sample (Gomez et al. 2016). We used perplexity to measure the applicability of the numbers of topics in the cross-validation (applicability of a topic model to new data). As shown in Figure 1, there was overfitting by the time we had 75 topics: the model began to deteriorate after 75. We can see that perplexity drops quickly below 10 topics. After 10, the speed is reduced, and it becomes relatively static after 50 topics. We wanted to examine closely the numbers before 50, so we used the second set (2, 3, 4, 5, 10, 16, 20, 30). As shown in Figure 1, the value dropped below 1,200 after 13 topics. Judgment is required to decide the number of topics we want from 13 to 50. Fifty has the lowest value, but it is too much for us to categorize all the topics in consumer research. Wang et al. (2015) fitted their model using 16 topics, and since 16 is inside our range [13, 50], we had random distributions starting at 16 topics in the end.

3, METHOD

3.1 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is a probabilistic topic modeling method. The intuition behind LDA is that a topic is assumed to be a certain distribution over words. We could model the documents and find the hidden distribution. We assume that there are some topics within a collection. Each document within this collection contains these topics with different probabilities (proportions). Out of the topics that the collection has, only a handful might be activated in each document (high probabilities of some topics and very low probabilities of others). Both word distribution and topic distribution are modeled by a Dirichlet distribution, with different parameters. For example, the topic distribution for each document is distributed as $\theta \sim \text{Dirichlet}(\alpha)$, and the word distribution is distributed as $\phi \sim \text{Dirichlet}(\eta)$. The values of both α and η denote the number of topics/words that one document/topic is likely to contain.

We chose Dirichlet because of its property of conjugating to the multinomial. The posterior distribution of θ, given this topic indicator, would still be at Dirichlet. This property would speed up and simplify our computation. We could use the LDA framework to learn the word distribution underlying each topic and the topic distribution in each document. It can be used to study the evolution of topics by calculating their probabilities over different periods within certain documents. LDA is, therefore, a powerful model that can be used to visualize the hidden thematic structure from a large corpus. As shown in Table 1, scholars use LDA to uncover topics and show trends from abstracts in different fields in academic journals.

3.2 Dynamic topic model

Traditional LDA assumes that the documents are interchangeable. The problem with this is that the same topic may change over time and that documents may therefore not be interchangeable within the corpus. In a large corpus, especially archives of document collections that have a huge time range, the words underlying each topic could change. People in 1974 may have used a whole different set of words than people in 2007 when talking about "advertising," "social identity," etc. If the topic was found using mostly words from recent decades, then the distribution of the same topic over the documents in the last decade could be underestimated because there may not be enough words to increase the probability. Hence, traditional LDA may not be able to catch the change and find the real distribution of topics on documents over time. When the documents are well organized in sequence, dynamic topic model (DTM) can be efficient in analyzing the corpus. We took logistic normal distribution (Aitchison 1980) and combined that with a state space model (West and Harrison 1997) to model topics evolving over time.

To link the topics in time slice, we used a generative process from Blei and Lafferty (2009)[2].

(1) Draw topics $\vec{\pi}_t \mid \vec{\pi}_{t-1} \sim N(\vec{\pi}_{t-1}, \sigma^2 I)$
(2) For each document:
  (a) Draw $\theta_d \sim \mathrm{Dir}(\vec{\alpha})$
  (b) For each word:
    (i) Draw $Z \sim \mathrm{Mult}(\theta_d)$
    (ii) Draw $W_{t,d,n} \sim \mathrm{Mult}(f(\vec{\pi}_{t,z}))$.

They treated each time slice as a separate LDA model and chained each topic to its predecessor and successor. Since DTM divides the data by time slice and adapts to the updated component of each topic, it does not miss the word changes underlying each topic over time and can therefore capture the evolution of topics efficiently.

Helgeson et al. (1984) reviewed and classified consumer behavior literature from ten top publications[3]. They examined the growth of consumer behavior topics through 15,000 articles from 1950 to 1981, grouping topics into four major areas: internal, external, purchase process, and miscellaneous. They found that the number of topics increased steadily over the years. Environmental changes (e.g., societal, technological, economic, public policy, etc.) have affected consumer behavior topics. As shown in their findings, topics such as preference, attitudes, and perception have appeared consistently in the literature since the 1950s. Among those topics found

[2] Generative process from Blei and Lafferty (2009).

[3] Journal of Consumer Research, Journal of Marketing, Journal of Marketing Research, Journal of Advertising, Journal of Advertising Research, Harvard Business Review, Journal of Business, Journal of Applied Psychology, Proceedings of the Association for Consumer Research, and Proceedings of the American Marketing Association

in their study, there are at least three characteristic life cycle patterns (maturity, growth, and learning). The current study reviewed JCR and ACR from 1974 through 2017 (44 years) with 14,286 articles and 1,455 unique words, looking to discover the hidden thematic structure in the collection and to investigate the evolution of topics over these four decades using DTM. Some findings in current study corresponded to the findings in Helgeson et al. (1984).

4, WORDS AND HIERARCHICAL CLUSTERS

Table 2 reports the top words used in JCR and Table 3 reports the top words used in ACR in four time slices. The top twenty words in the first time slice in JCR are "consum," "product," "pruchas," "time," "differ," "cognit," "prefer," "famili," "group," "social," "process," "respons," "strategi," "level," "structur,""base," "consumpt," "price," and "altern." The top twenty words in the first time slice in ACR are "consum," "product," "subject," "time," "group," "level," "respons," "price," "differ," "purchas," "involv," "social," "person," "factor," "tabl," "categori," "base," "question," and "adverti." Some of these top twenty words fade away in later time slices. However, as shown in the tables, the top words in the four time slices for both JCR and ACR align with common practical frameworks.

Among those top words, some have always been popular, such as "consum," "price," "product," "goal," "feel," and "adverti." These are leading words in JCR as well as ACR and remain popular across the decades. Some words in the lists fade in growth and some eventually die away. In JCR, "social" became increasingly popular in 2007–2017. The word "cultur" became very popular in the 1996–2006 time slice, but its use declined in 2007–2017 as "social" became popular. The word "differ" died down in the last period. Similarly, the use of "social" also increased in ACR in the last decade. The same changes in both JCR and ACR might be associated with environmental changes (societal, technological, economic, public policy, etc.) in this era.

The popularity of the same words differs across journals. As shown in the tables, the common word rank (number) in each journal is different for each journal. For example, the word that indicates self-control has different popularity in JCR and ACR. It ranks sixth in the third time slice in ACR, but does not appears in JCR until the last time slice (2007–2017).

Although the frequency lists only show the top 20 words because of space limitations, Figure 2 reports the word clouds of abstracts published in JCR and ACR in four time slices. To further see how the terms are associated with each other, we perform a cluster analysis using corlimit[4]:

> *The corlimit is a numeric vector for the lower correlation limits of each term (ranges 0:1). Each list component is named after a term in terms and contains a named numeric vector. Each vector holds matching terms from document term matrix (dtm) and their rounded correlations satisfying the inclusive lower correlation limit of corlimit.*

As shown in Table 4, the four-cluster dendrograms denote the hierarchical cluster analysis of abstracts published in JCR. JCR_ab01 represents the abstracts published in JCR from the first time slice (1974–1984). The size of the cluster increased from the first time slice to the last one. In the first time slice, "product" is highly associated with "time," "differ," and "purchas" only. In the second time slice, "product" is highly associated with more words. This trend continues in the third and last time slice. This may indicate that more terms/topics have evolved in articles published in JCR. The connections between topics are expanding. In the cluster dendrogram, the word "social" could be found in the last two time slices, corresponding to the word and topic trends

---

[4] Corlimit defined in package “tm” in R

shown in Figure 4 and Figure 5. We believe that the topics associated with word "social" could have the "growth" and "learning" life cycle patterns found in Helgeson et al. (1984). The next section shows how topic and word use evolve over time.

5, TOPIC EVOLUTION AND TRENDS OF WORDS

There are words like buyer, buy, buying, purchase, shopping, child, and children in the top twenty words list underlying the topics in Wang et al. (2015). We post-stemmed the list of words and saved the top space for diverse tokens.

As shown in Table 6, each of the topics is a distribution over words, and the words underlying each topic were ordered by their probability. Those words at the top of the list (with high probability) correspond to things that we recognize. Thus, though those topics do not have names, we can name them based on words with high probabilities underlying each topic. Similar to the findings of Wang et al. (2015), we found the topics of advertising, buying process, self-control, family decision-making, memory price and price associations, resource constraints, social identity and influence, and methodological issues. Most of the topics are not activated on each document. As shown in Table 5, out of 16 topics, only a handful seems to have been activated on each article. The distribution over word and topic has been further examined using a dynamic topic model.

Figure 3 shows the dynamic topic model in Blei (2009). As shown in the figure, each slice is the same LDA model, where each document within each slice is exchangeable. The documents are not exchangeable across slices. To find the topic evolution, we divided our corpus into four sequential slices by years in both JCR and ACR (2×4), and we combined the time slices of the two publications (1×4).

We first took all the articles together and fitted models with random distributions of topics over words that generated those articles. Then we used the four sequential slices divided earlier to examine the evolution using dynamic topic modeling. There are distributions over words through each time slice. We treated each of the time slices as a separate LDA model. There are 16 topics for each slice, and we chained each topic to its predecessor and successor (as shown in Figure 3). Since DTM divides the data by time slice and adapts to the updated component of each topic, it would not miss word changes underlying each topic.

The model seems to have captured something salient about these abstracts through time. We noticed that words underlying topics have a change in probability over time. As shown in Figure 4, the topic "advertising" marched forward decades and changed a little bit in each time slice. The four columns in Figure 4 show the change in the top word component underlying the same topic "advertising" from 1974–1984 and up to 2007–2017. The way advertising was talked about in 1974–1984 seems to be different from the way it was talked about in 2007–2017. In the topic "advertising," words like "communic", "risk", etc almost faded out over time, and words like "messag" grow through the time slices. As we mentioned earlier, "social" has become popular in the last decades. We can also identify this growing trend under the topic "advertising." A shown in the figure, the probability of "social" has been increasing since 1996–2006.

We further used DTM to investigate the topic distributions over abstracts from 1974 through 2017. Figure 5 reports the trend of "family decision making," "search for information," "advertising," "socialization," "choice," and "price." It seems that "socialization" and "search for information" have a growth pattern. The distributions of "family decision making" and "choice" are static over time. "Price" seems to have a maturity pattern. These distributions corresponded to the findings in Helgeson et al. (1984) but not to the findings in Wang et al. (2015). Next, we wanted to further use those hidden themes captured by DTM to conduct a similarity assessment between publications.

6, INTERACTION BETWEEN JCR and ACR

The classical similarity metrics will not find any similarity between articles because classical similarity metrics are based only on word use, which can drastically change over time. Hellinger distance was introduced by Ernst Hellinger (1909) and is used to measure the similarity between two probability distributions (either discrete or continuous). This is like a time corrected document similarity metric. The symmetric distance between distributions can be used to measure how similar two articles are. In current study, the similarity is defined by the similarity between our document proportions. In Hellinger distance, θ is a hidden variable. We took its expectation, given the different documents. We want to see how similar the θ are to each other after taking the θ for one document and the θ for another document.

The similarities could be scored and ranked using distance measurement, and cross the entire sample of 14,286 articles published in the two publichations, we could find a similar article to one article. After combing all the articles published in each year into one slice (2x44 slices), we tested the similarities across publications, without worrying about the word changes. We found that JCR and ACR tend to have the smallest distance when they are close to each other in time. Based on this, we wanted to see if topics in ACR would have some impacts on the JCR topics in the next year. As shown in Table 5, each topic has a distribution over each document. We used this to calculate the topic distribution over documents published in JCR and ACR each year (treated each year as one document). Abstracts published within one year were combined into one document in both publications (1×88). Then we treated the topic distribution as variables:

$$JCR_{1,t} = \beta_{1,0} + \beta_{1,1} ACR_{1,t-1} + \varepsilon_1$$

$$JCR_{2,t} = \beta_{2,0} + \beta_{2,1} ACR_{2,t-1} + \varepsilon_2$$

$$\vdots$$

$$JCR_{16,t} = \beta_{16,0} + \beta_{16,1} ACR_{16,t-1} + \varepsilon_{16}$$

However, we could not find any significant parameters. Using time corrected similarity metrics, we found an interaction between the two publications when they are close in time. But we could not identify a particular topic impact over the same topic in the two publications. The time corrected similarity metric takes into account the structure everywhere to calculate the similarity among documents. Hence, we conjectured that a single topic might fail to have a powerful influence on that same topic across publications.

7, DISCUSSION and CONCLUSIONS

We examined the hidden thematic structure in the observed words in 14,286 articles published in JCR and ACR. We further investigated the evolution of substantive topics over the years. We studied how the development of topics in Advances in Consumer Research Proceedings (ACR) might relate to the growth in topics at Journal of Consumer Research (JCR). The current study used DTM to capture the word change underlying each topic. The analysis generated a number of insights. First, both JCR and ACR top words in four time slices align with common practical frameworks. The figures in our findings suggest that some words have always been popular, some are in growth, and some die away. Some changes occurred in both JCR and ACR. The use of "social" increased in ACR in the last decade. In JCR, "social" became increasingly popular in 2007–2017. The same changes in both publications might be associated with environmental change. Second, we noticed that words underlying topics have probability changes over time. The way a topic was talked about a few decades ago differs from how topics are talked about now. Third, although we found that some topics correspond to the findings of Wang et al. (2015), their distributions were not the same. Furthermore, the interaction between the two publications, JCR and ACR, tend to have the smallest distance when they are close to each other in time, whereas a single topic might fail to have a powerful influence on the same topic across these two publications. Future research could seek to better understand how trends in conferences drive trends in research and whether applied topics lead to research or research leads to applied topics.

8, REFERENCES

Baumgartner H & Pieters R (2003). The structural influence of marketing journals: A citation analysis of the discipline and its subareas over time. Journal of Marketing, 67(2):123-139.

Baumgartner, H (2010). Bibliometric reflections on the history of consumer research. Journal of Consumer Psychology, 20(3): 233-238.

Blei DM & Laff erty JD (2009). Topic models. Text Mining. Chapman and Hall/CRC, 101-124.

Buzzell RD (1968). Mathematical Models and Marketing Management. Harvard University Press, Cambridge, UK.

Hall CM (2006). The impact of tourism knowledge: Google scholar, citations and the opening up of academic space. e-Review of Tourism Research, 4(5): 119-136.

Hall CM (2011). Publish and perish? Bibliometric analysis, journal ranking and the assessment of research quality in tourism. Tourism Management, 32(1): 16-27.

Helgeson JG, Kluge EA, Mager J, & Taylor C (1984). Trends in consumer behavior literature: A content analysis. Journal of Consumer Research, 10(4): 449-54.

Howey RM, Savage KS, Verbeeten MJ, & Van Hoof HB (1999). Tourism and hospitality research journals: Cross-citations among research communities. Tourism Management, 20(1): 133-139.

Jamal T, Smith B, & Watson E (2008). Ranking, rating and scoring of tourism journals: Interdisciplinary challenges and innovations. Tourism Management, 29(1): 66-78.

Kevork EK & Vrechopoulos AP (2009). CRM literature: Conceptual and functional insights by keyword analysis. Marketing Intelligence and Planning, 1(1): 48-55.

Leeflang PSH, Wittink DR, Wedel M, & Naert PA (2000). Building Models for Marketing Decisions (Kluwer Academic Publishers, Boston).

Lilien GL (1994). Marketing models: Past, present and future. Laurent G, Lilien GL, & Pras B, ed. Research Traditions in Marketing (Kluwer Academic Publishers, Boston), 1-26.

Lilien GL, Kotler P, & Moorthy KS (1992). Marketing Models (Prentice-Hall, Englewood Cliffs, NJ).

McKercher B (2005). A case for ranking tourism journals. Tourism Management, 26(5): 649-651.

McKercher B (2008). A citation analysis of tourism scholars. Tourism Management, 29(6): 1226-1232.

McKercher B, Law R, & Lam T (2006). Rating tourism and hospitality journals. Tourism Management, 27(6): 1235-1252.

Mela CF, Roos J & Deng YT (2013). A keyword history of marketing science. Marketing Science 32(1): 8-18.

Moussa S & Touzani M (2010). Ranking marketing journals using the Google scholar-based hg-index. Journal of Informetrics, 4(1):107-111.

Nocosia, FM. 1969. Consumer research: Problems and perspectives. Journal of Consumer Affairs, 3(1): 9-25.

Ryan C (2005). The ranking and rating of academics and journals in tourism research. Tourism Management, 26(6): 657-662.

Schofield A & Mimno D (2016). Comparing apples to apple: The effects of stemmers on topic models. Transactions of the Association for Computational Linguistics, 4: 287-300.

Tellis GJ, Chandy RK & Ackerman DS (1999). In search of diversity: The record of major marketing journals. Journal of Marketing Research, 36(1): 120-131.

Wang X, Bendle NT, Mai F & Cotte J (2015). The Journal of Consumer Research at 40: A historical analysis. Journal of Consumer Research, 42(1): 5-18.

9, FIGURES and TABLES

Table 1 Studies

Table 1.1 Example Studies of Topic Models of Academic Papers

Year	Author(s)	Publication Title	Data	Model
2013	Mela et al.	Marketing Science	1,050 articles	LDA
2017	JungSu et al.	J Counseling Psychology	3,603 articles	LDA
2015	Xin et al.	J Consumer Research	40 years	LDA
2010	Rosen-Zvi et al.	ACM Trans on Info Sys	Abstracts, papers	Two-stage stochastic process
2017	Choi et al.	Computers & Security	2,356 articles	LDA
2016	Mortenson et al.	Int J of Info Mgmt	3,386 articles	Technology acceptance model
2008	Hall, et al.	EMNLP'08	12,500 articles	LDA
2006	Blei et al.	23rd Int Conf on Machine Learning	30,000 articles	DTM

Table 1.2 Studies in Consumer Research History

Year	Author(s)	Publication Title
2015	Xin et al.	J Consumer Research
2014	Troung et al.	Social Marketing Quarterly
2013	Logemann et al.	Business History Review
1969	Nicosia	J Consumer Affairs
1984	Helgeson et al.	J Consumer Research
2010	Baumgartner et al.	J Consumer Psychology

Table 2 Most Frequently Used Words in Journal of Consumer Research[5]

1974-1984	**freq**	1985-1995	**freq**	1996-2006	**freq**	2007-2017	**freq**
consum	243	consum	569	consum	662	consum	1545
product	88	product	245	price	210	product	504
purchas	63	price	218	product	205	consumpt	371
time	60	process	120	process	156	goal	328
differ	50	respons	110	affect	145	experi	274
cognit	48	subject	110	prefer	120	social	259
prefer	46	affect	107	cultur	113	prefer	251
famili	45	purchas	106	base	111	time	224
group	44	level	101	goal	109	affect	206
social	44	consumpt	100	consumpt	105	level	202
process	43	base	95	differ	104	ident	196
respons	42	altern	94	altern	102	control	192
strategi	42	differ	88	context	102	price	180
level	41	condit	82	time	94	lead	175
structur	40	experi	80	social	91	emot	168
base	39	social	76	purchas	90	context	161
consumpt	39	search	74	impact	89	purchas	160
price	36	memori	72	level	87	activ	156
altern	35	time	70	persuas	86	option	155

[5] Show the top 20 words due to limited space

Table 3 Most Frequently Used Words in Advances in Consumer Research

1974-1984	**freq**	1985-1995	**freq**	1996-2006	**freq**	2007-2017	**freq**
consum	60502	consum	61266	consum	49286	consum	47943
product	34462	product	21717	product	17552	condit	17755
subject	19502	time	17830	condit	17220	social	15292
time	19193	price	13565	goal	16331	product	13766
group	18099	consumpt	12870	consumpt	13753	experi	11960
level	17562	cultur	12566	control	13084	consumpt	11229
respons	16470	experi	12004	time	11359	time	11033
price	16386	person	11833	experi	11296	control	10860
differ	15719	journal	11628	emot	11202	prefer	10222
purchas	15701	purchas	10784	social	9726	purchas	9387
involv	15611	goal	10396	level	9044	food	9259
social	15097	level	9836	person	8836	goal	8356
person	13799	social	9819	prefer	8272	feel	7825
factor	13612	condit	9606	ident	7991	person	6861
tabl	13550	group	9566	feel	7901	emot	6589
categori	11430	base	9300	purchas	7808	impact	6333
base	10841	emot	8810	task	7599	task	6189
question	10642	differ	8504	price	7436	affectt	6173
adverti	10573	affect	8226	cultur	7381	percept	6018

Table 4 Hierarchical Cluster Analysis in Journal of Consumer Research

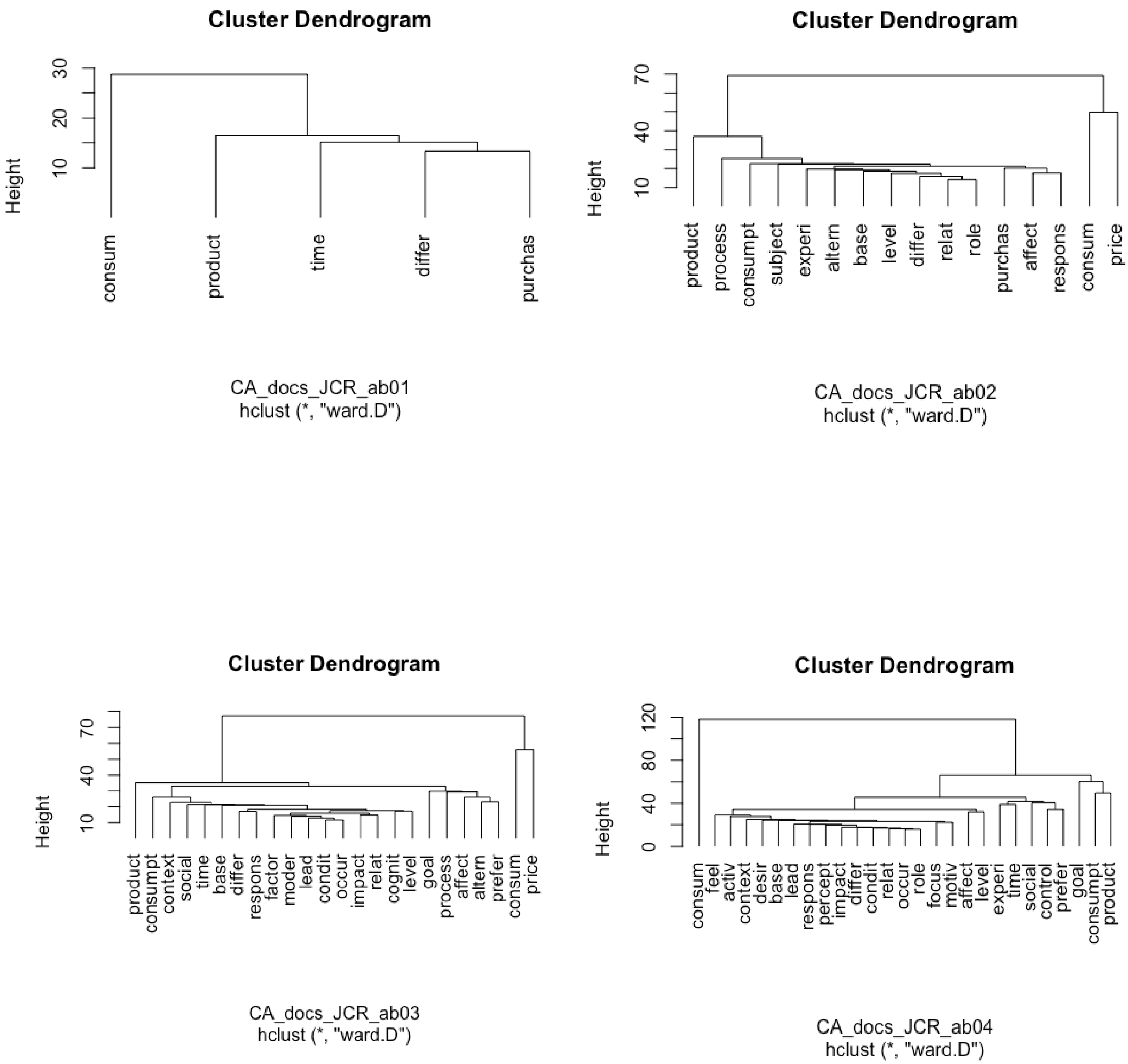


Table 5 Topic Distribution in Document[6]

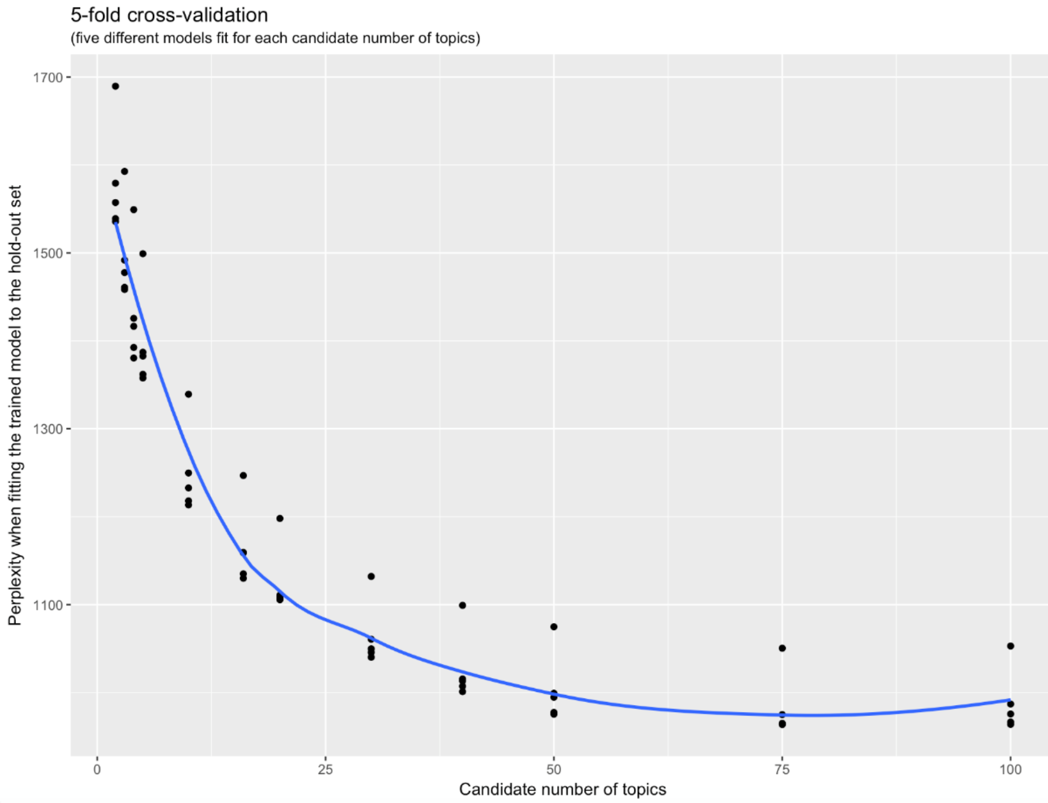
Document	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11	Topic 12	Topic 13	Topic 14	Topic 15	Topic 16
1	0.0473	0.1080	0.0473	0.0473	0.1686	0.0473	0.0473	0.0625	0.0473	0.0473	0.0473	0.0473	0.0473	0.0625	0.0473	0.0777
2	0.0581	0.0581	0.0440	0.0722	0.0440	0.0440	0.1004	0.0440	0.0581	0.0440	0.1144	0.0581	0.0722	0.0863	0.0581	0.0440
3	0.0712	0.0990	0.0712	0.0434	0.0434	0.0712	0.0573	0.0573	0.0712	0.0712	0.0434	0.0434	0.0573	0.0573	0.0851	0.0573
4	0.1195	0.0460	0.0754	0.0460	0.0607	0.0607	0.0607	0.0460	0.0460	0.0607	0.1048	0.0460	0.0460	0.0460	0.0901	0.0460
5	0.0539	0.0711	0.0539	0.0711	0.0539	0.0539	0.0539	0.0711	0.0539	0.0711	0.0539	0.0539	0.0884	0.0711	0.0539	0.0711
6	0.0488	0.0645	0.0957	0.0957	0.0645	0.0645	0.0488	0.0645	0.0488	0.0488	0.0645	0.0488	0.0801	0.0645	0.0488	0.0488
7	0.0488	0.0488	0.0645	0.0801	0.0645	0.1113	0.0801	0.0645	0.0645	0.0645	0.0488	0.0488	0.0488	0.0488	0.0488	0.0645
8	0.0466	0.0466	0.1063	0.0466	0.0616	0.0616	0.0616	0.0466	0.0466	0.0466	0.0616	0.1362	0.0466	0.0616	0.0616	0.0616
9	0.0581	0.1285	0.0440	0.0440	0.0440	0.0440	0.0440	0.0440	0.0440	0.0440	0.0440	0.0440	0.1144	0.0581	0.0863	0.1144
10	0.0496	0.0496	0.0972	0.0972	0.0496	0.0655	0.0496	0.0496	0.0655	0.0655	0.0496	0.0813	0.0813	0.0496	0.0496	0.0496

Table 6 The Representativeness of Terms Within 16 Topics [7]

	Satisfying Customers	Choice	Family Decision Making	Evaluation	Self-control	Methodological Issues	Social Identity and Influence	Socialization
1	goal	product	pattern	intent	consum	consum	social	emot
2	option	prefer	relat	structur	consumpt	critic	ident	activ
3	motiv	level	famili	behavior	control	dimens	group	person
4	situat	categori	gift	construct	focus	design	materi	object
5	satisfact	featur	role	valid	food	address	possess	role
6	achiev	construal	household	util	desir	demand	status	lead
7	lead	benefit	econom	procedur	health	introduc	relationship	contribut
8	impact	differenti	women	estim	promot	fail	share	prime
9	action	origin	work	multipl	regulatori	psycholog	express	work
10	pursuit	countri	adopt	design	opportun	approach	form	play
	Buying Process	**Age different**	**Advertising**	**Price**	**Search for Information**	**Resource Constraints**	**Memory**	**Belief-Expectancy**
1	altern	differ	affect	price	consumpt	time	process	experi
2	context	cue	advertis	purchas	cultur	feel	memori	belief
3	perform	subject	condit	qualiti	practic	futur	base	infer
4	factor	strategi	cognit	refer	compar	resourc	attitud	expect
5	risk	children	messag	search	info	power	exposur	servic
6	percept	bias	persuas	cost	emerg	money	visual	event
7	comparison	relat	sourc	offer	explor	distanc	attent	commerci
8	frame	age	impact	discount	ideolog	reward	recal	occur
9	attract	adult	communic	percept	inform	spend	learn	cost
10	reason	similar	involv	buy	symbol	receiv	judgment	hedon

[6] 10 articles from JCR 1974-1984
[7] Show top 10 words in the model; Topics resulted from all articles

Figure 1 5 Fold Cross-Validation



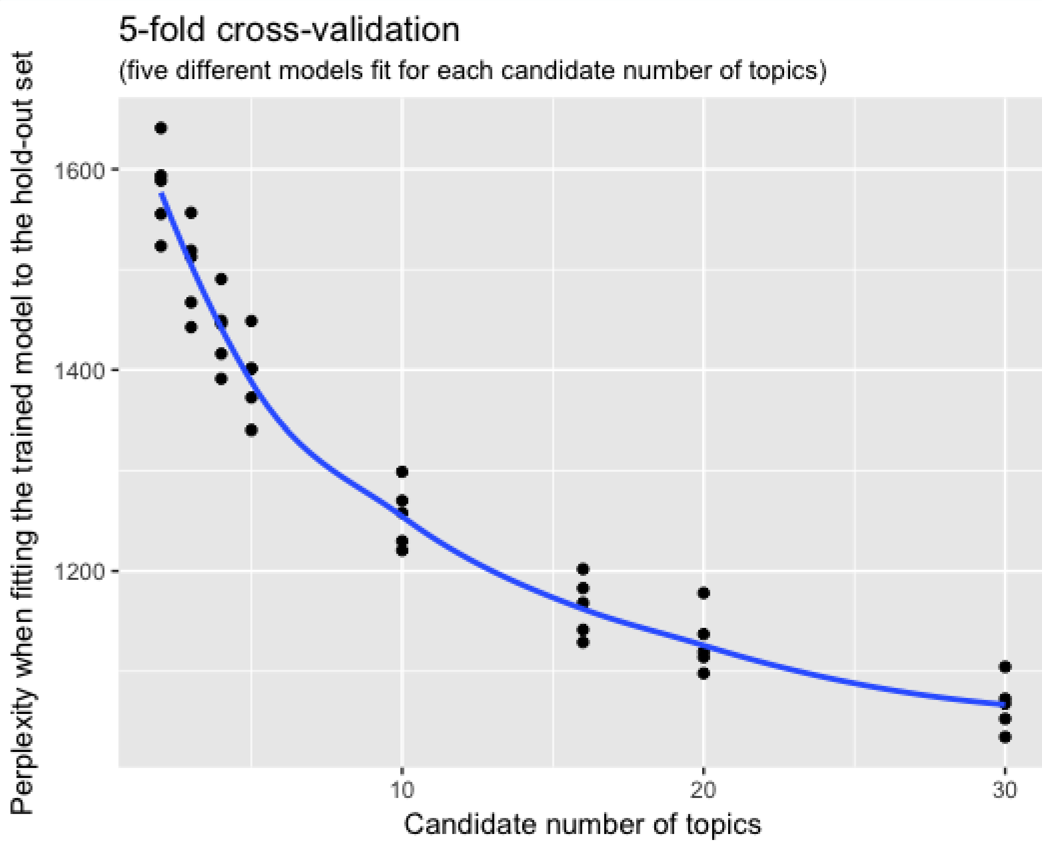


Figure 2 Word Cloud by Decades in JCR and ACR (4)

1,

2,


price
product
consumpt
process
subject
affect
respons
search
altern
differ
categori
social
experi
qualiti
purchas
level
memori
structur
condit
context
advertis
cognit
children
role
time
base
relat
recal
impact
involv
messag
attitud

consum

3,


consum
product
price
affect
prefer
differ
altern
goal
process
consumpt
purchas
option
cultur
impact
base
social
messag
cognit
experi
categori
level
condit
refer
context
memori
persuas
respons
time
factor
emot
qualiti
search
resourc
consider
featur
relat
cue
object
task
attract
risk
benefit
bias
focus
percept
intent
pattern
judgment
reason
moder
similar
elabor
thought
promot
perform
occur
group
chang
retriev
outcom
form
relationship
lead
strategi
advertis
cost
attitud
sourc
structur
subject
construct
frame
comparison
regret
learn
motiv
visual
offer
recal
possess
mood
agent
role
appeal
assess
estim
person

4,


consum
product
consumpt
goal
social
prefer
experi
time
emot
ident
price
option
affect
level
control
lead
context
cultur
purchas
percept
impact
feel
process
food
resourc
power
activ
construal
group
desir
motiv

Figure 3 Dynamic topic model[8]

22 D. M. BLEI AND J. D. LAFFERTY

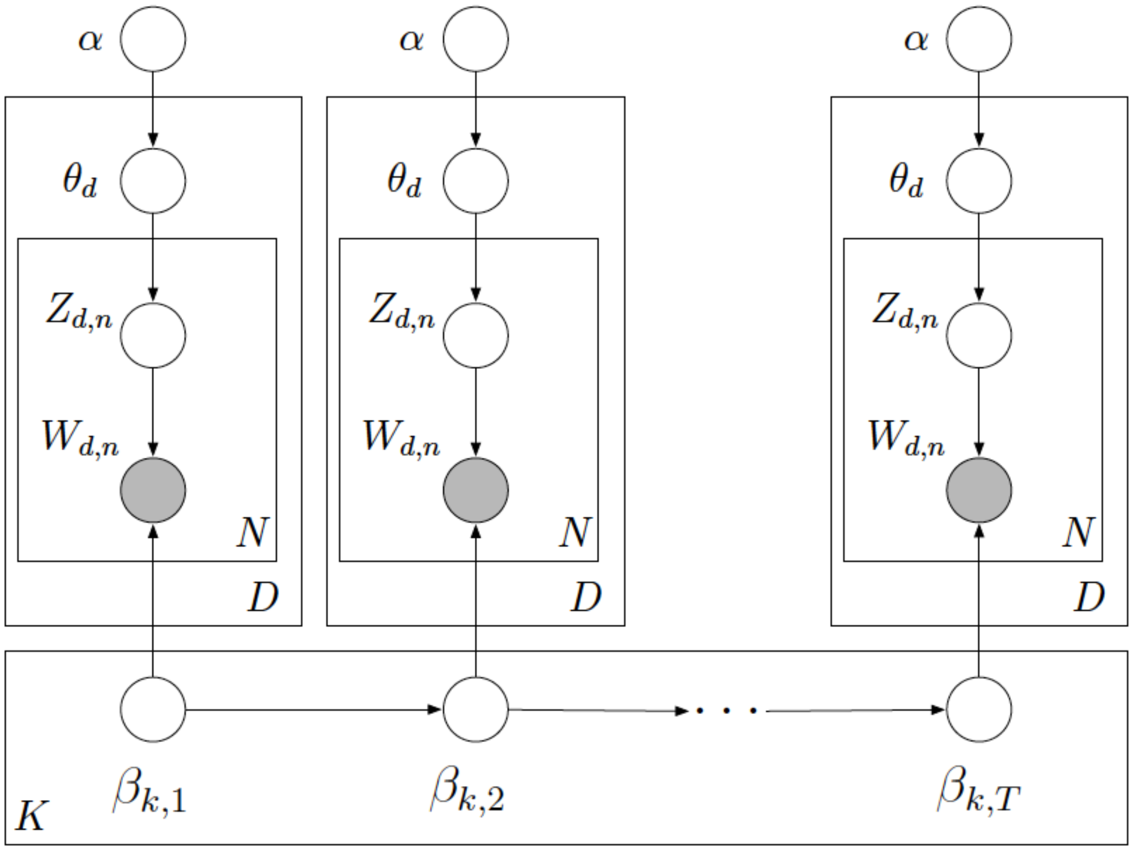


FIGURE 8. A graphical model representation of a dynamic topic model (for three time slices). Each topic's parameters $\beta_{t,k}$ evolve over time.

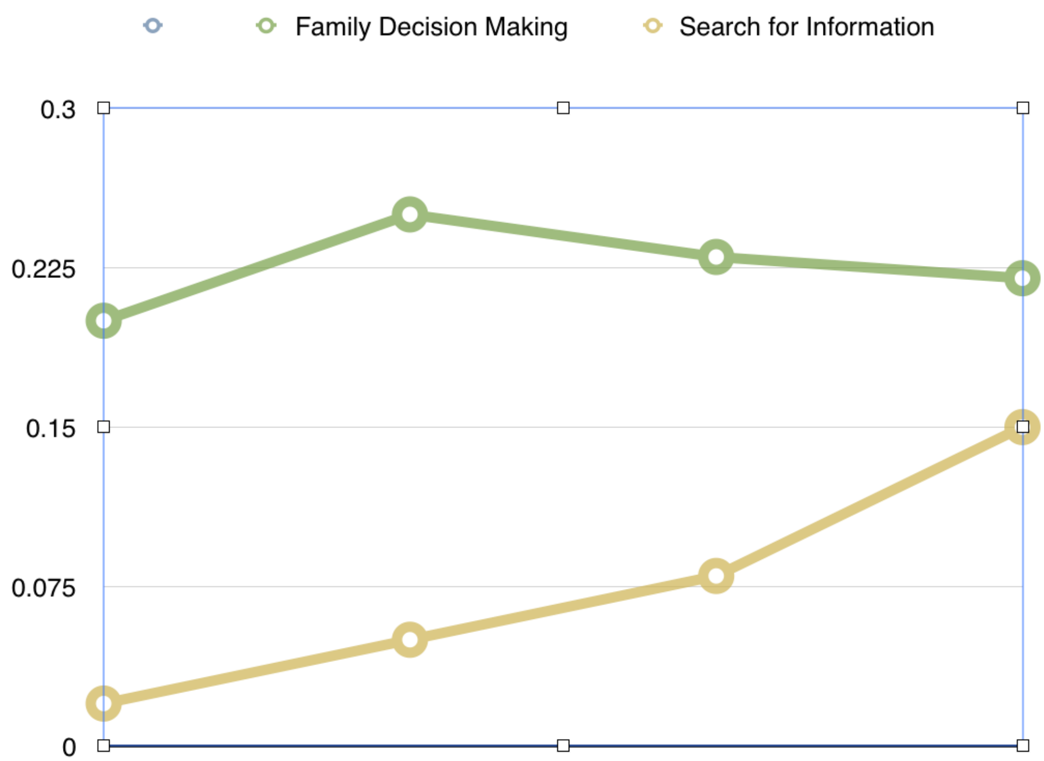
[8] Blei, Lafferty (2009)

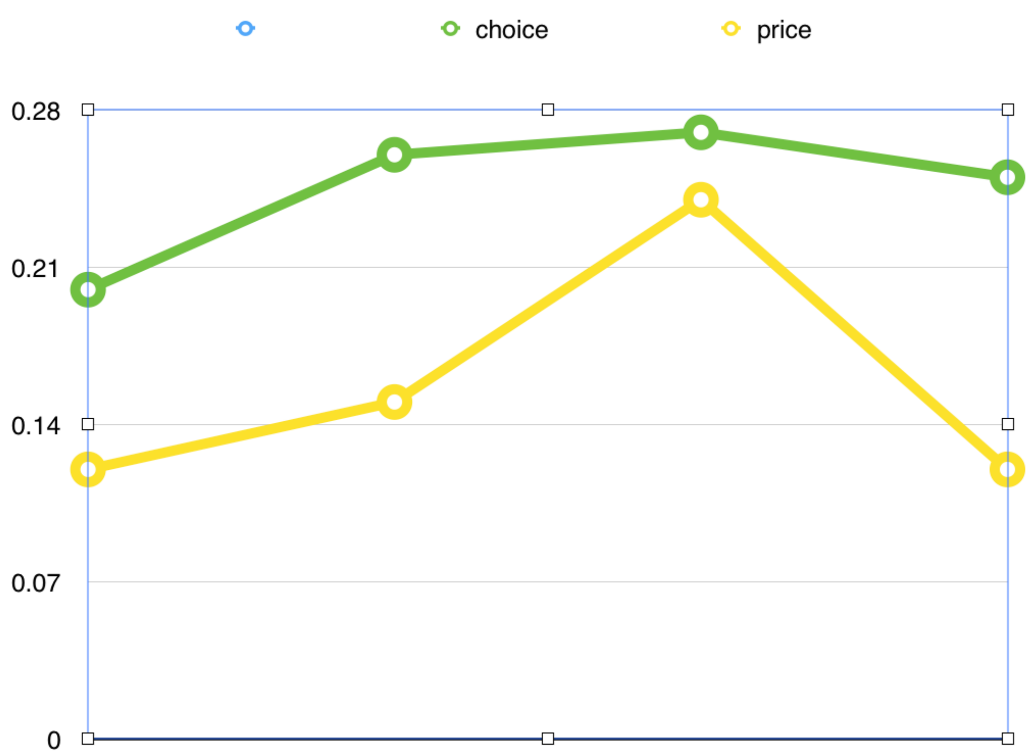
Figure 4 Topic Evolution and Trend of Word Within Topic

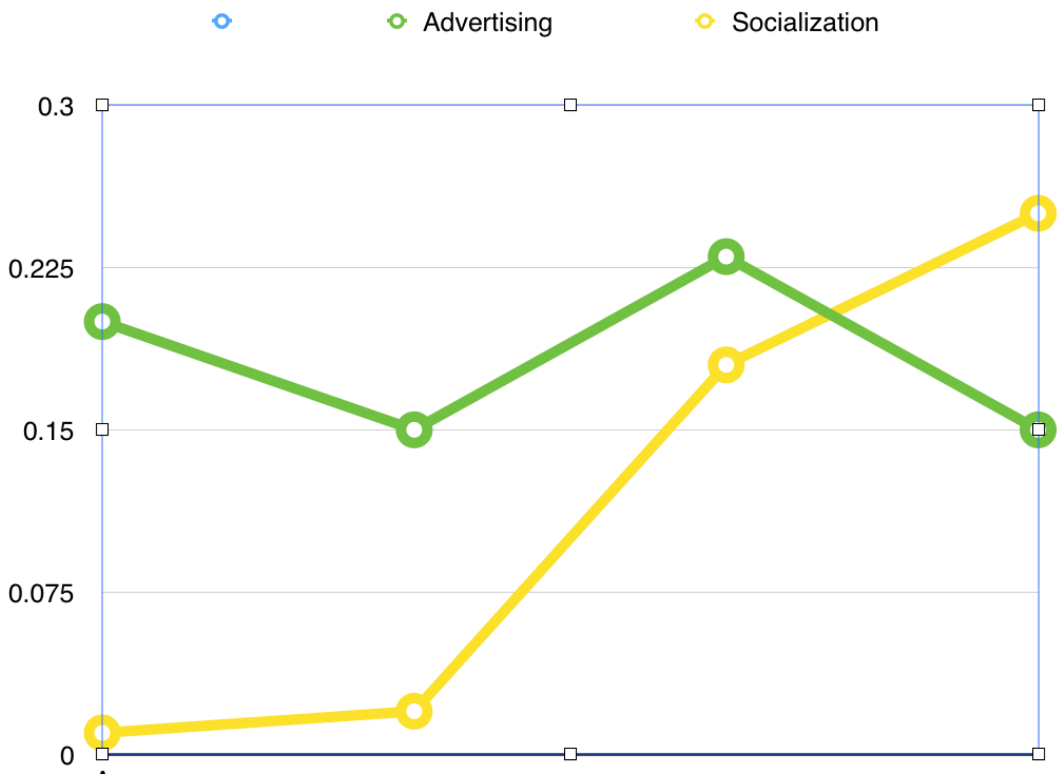
Examples of Advertising

1974-1984	1985-1995	1996-2006	2007-2017
communic	affect	context	focus
advertis	risk	attitude	persuas
condit	advertis	involv	messag
affect	relat	cognit	lead
sourc	replic	messag	cognit
persuas	distinct	persuas	custom
messag	sourc	sourc	involv
impact	persuas	impact	social
communic	communic	communic	advertis
involv	messag	social	communic

Figure 5 Topic Trends
Examples: Six Topics in JCR (1974-1984, 1985-1995, 1996-2006, 2007-2017)






Advertising
Socialization
0.3
0.225
0.15
0.075
0