



## Management Science

Publication details, including instructions for authors and subscription information:  
<http://pubsonline.informs.org>

### Content-Based Model of Web Search Behavior: An Application to TV Show Search

Jia Liu , Olivier Toubia , Shawndra Hill

To cite this article:

Jia Liu , Olivier Toubia , Shawndra Hill (2021) Content-Based Model of Web Search Behavior: An Application to TV Show Search. Management Science 67(10):6378-6398. <https://doi.org/10.1287/mnsc.2020.3827>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact [permissions@informs.org](mailto:permissions@informs.org).

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2021, INFORMS

Please scroll down for article—it is on subsequent pages






With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

# Content-Based Model of Web Search Behavior: An Application to TV Show Search

Jia Liu,<sup>a</sup> Olivier Toubia,<sup>b</sup> Shawndra Hill<sup>c</sup>

<sup>a</sup> Department of Marketing, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong; <sup>b</sup> Graduate School of Business, Columbia University, New York, New York 10027; <sup>c</sup> Wharton Customer Analytics, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Contact: jiali@ust.hk,  <https://orcid.org/0000-0002-0279-724X> (JL); ot2107@columbia.edu,  <https://orcid.org/0000-0001-7493-9641> (OT); shawndra@wharton.upenn.edu,  <https://orcid.org/0000-0003-1980-727X> (SH)

Received: June 15, 2018

Revised: March 28, 2019; February 4, 2020;  
July 20, 2020

Accepted: August 7, 2020

Published Online in Articles in Advance:  
February 26, 2021

<https://doi.org/10.1287/mnsc.2020.3827>

Copyright: © 2021 INFORMS

**Abstract.** We develop a flexible content-based search model that links the content preferences of search engine users to query search volume and click-through rates, while allowing content preferences to vary systematically based on the context of a search. Content preferences are defined over latent topics that describe the content of search queries and search result descriptions. Compared with existing applications of topic modeling in marketing and recommendation systems, our proposed approach can simultaneously capture multiple types of information and investigate multiple aspects of behavioral dynamics in a single framework that enables interpretable results for business decision making. To facilitate efficient and scalable inference, we develop a full Bayesian variational inference algorithm. We evaluate our modeling framework using real-world search data for TV shows from the Bing search engine. We illustrate how our model can quantify the content preferences associated with each query and how these preferences vary systematically based on whether the query is observed before, during, or after a TV show is aired. We also show that our model can help the search engine improve its ranking of search results as well as address the cold-start problem for new page links.

**History:** Accepted by Hamid Nazerzadeh, big data analytics.

**Supplemental Material:** Data and the online appendices are available at <https://doi.org/10.1287/mnsc.2020.3844>.

**Keywords:** marketing • search • interpretable machine learning • recommendation systems • Poisson factorization • variational inference • big data

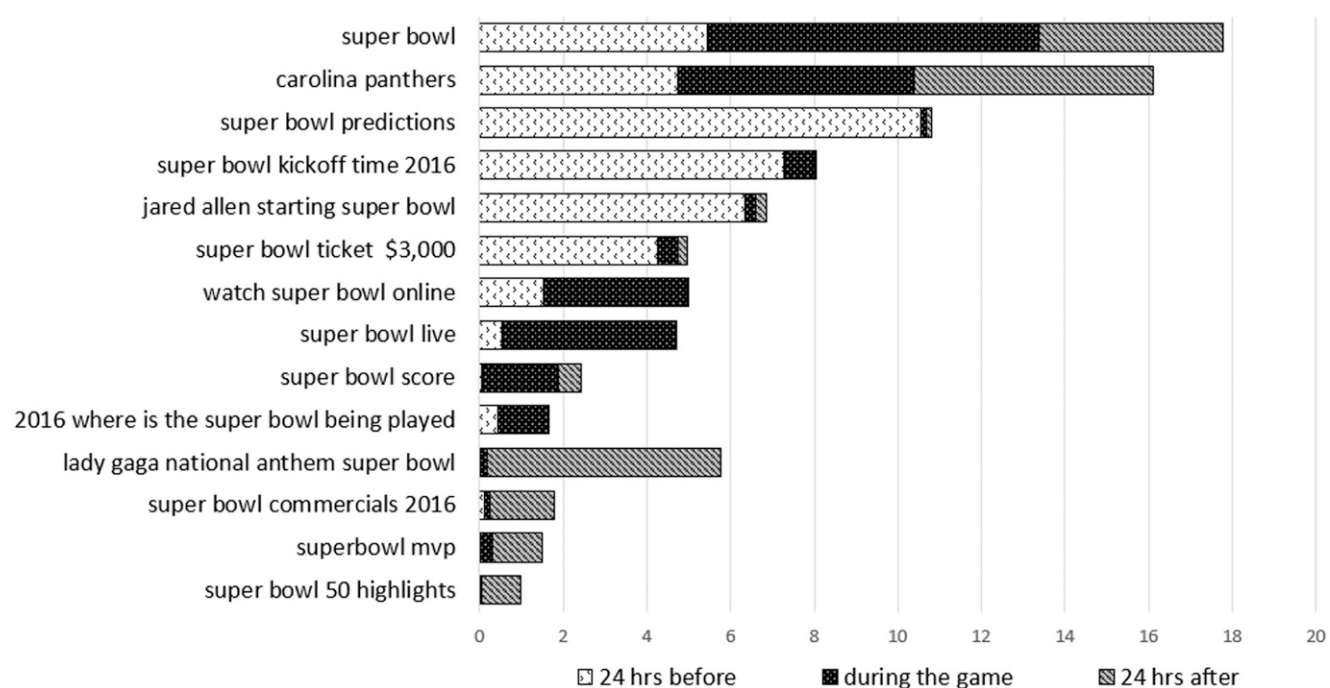
## 1. Introduction

Every minute, more than three million queries are entered in the Google search engine alone (Internet-LiveStats 2016), followed by clicks on the presented sponsored and organic results. Data on queries, page results, and subsequent clicks can provide tremendous insights into what type of information people seek over the course of a day. Based on a multitude of variables such as weather, daily and special events, time of day, and, in the case of this paper, TV events/shows that are aired, people's search intentions change. For example, Figure 1 reports the normalized total search volume for a few of the most popular queries by United States Bing users related to the 2016 Super Bowl, divided into three time windows: 24 hours before the game, the 4.5 hours during the game, and 24 hours after the game.<sup>1</sup> From the figure, it is apparent that some queries are used much more frequently than others, suggesting different overall levels of consumer interest across Super Bowl topics. In addition, some queries, such as "super bowl" and "carolina panthers," are searched for consistently over

time, whereas other queries exhibit usage dynamics. For example, before the game started, users tended to search for "tickets," "kickoff time," and "prediction"; during the game, users tended to search for "watch" and "live stream"; and after the game, users tended to search for "MVP," "commercial," and "highlights." Such variations in content preferences may be identified by simply observing variations in search volume over time.

Other variations in content preferences, however, are more subtle and can only be identified by observing downstream behavior such as clicks on returned links. Depending on when a query is entered (in this case, in relation to a TV event), the information that users wish to obtain by entering the query may vary, holding the query constant. These dynamics mean that users may click more frequently on some returned URLs based on the time they enter a search. As an illustration, Figure 2 displays the top 10 most frequently presented URLs on the search engine results page (SERP) for the query "super bowl," ranked by their average positions across impressions. A search

**Figure 1.** Normalized Search Volume for Super Bowl 2016 on Bing

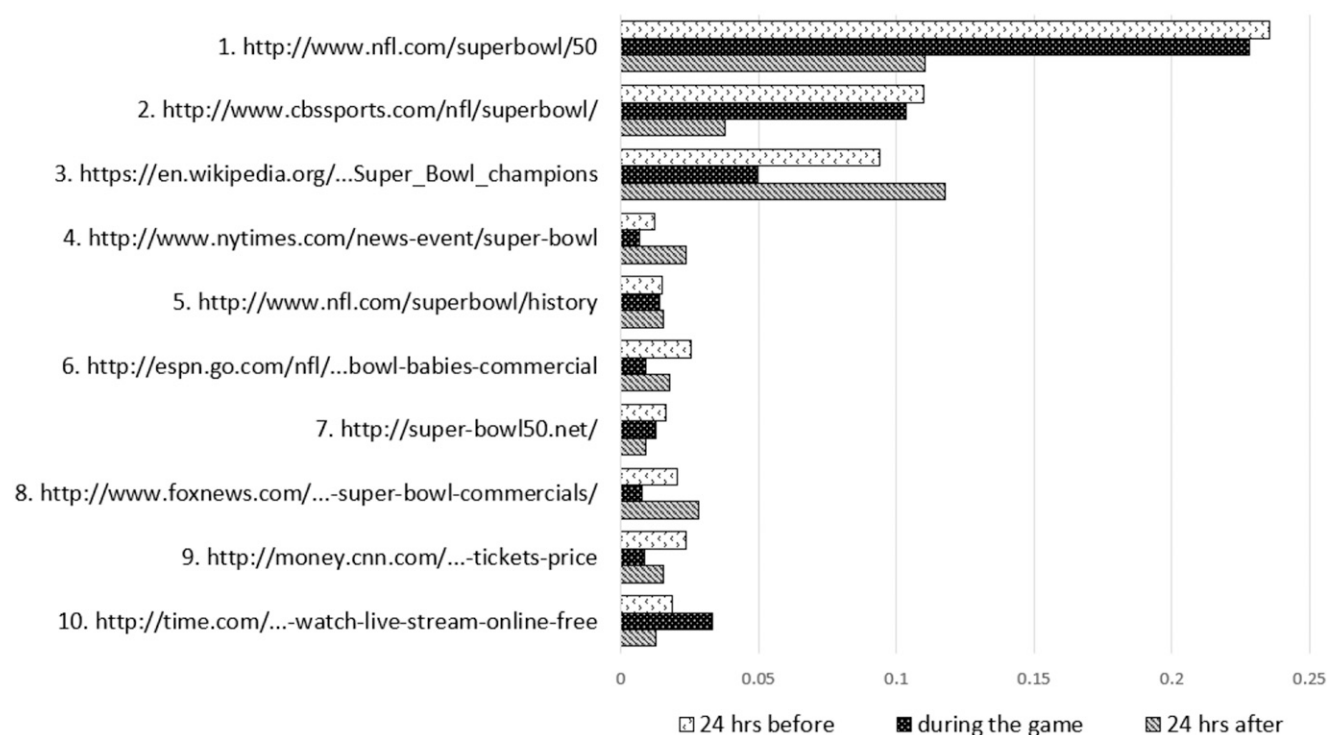


*Note.* For each sample query, this figure shows the proportion of searches issued during each of the three time windows (24 hours before, during, and 24 hours after the game was aired).

impression is defined as one set of organic search results returned by the search engine in response to one individual search.<sup>2</sup> Figure 2 reports the click-

through rate (CTR) for each URL within each time window, holding the query “super bowl” constant. As expected, there is generally a decrease in the CTR

**Figure 2.** CTRs of Top Search Results for “Super Bowl” on Bing



*Note.* The CTR of each link in each time window is computed as the number of observed clicks divided by the number of impressions.

as the position decreases. More interestingly, we also see significant differences in the CTR for several URLs across time. For example, the CTR for the third link, which is for Super Bowl champions on wikipedia.org, is 50% higher before the game than during the game, and the CTR for the fourth link, which is for Super Bowl news on nytimes.com, is at least 100% higher after the game than before and during the game. Importantly, when comparing the CTR across URLs within the same time window, it is possible to suggest a different ranking from that shown in Figure 2. For instance, after the game, the third link (from wikipedia.org about Super Bowl champions) should probably be ranked higher than the second link (from cbssports.com); during the game, the 10th link (about the channel airing the Super Bowl) should probably be ranked much higher. In other words, the positions of links on the SERP are not always consistent with the actual relevance of the links to users at a particular point in time, as measured by the observed CTR.

This example suggests that the search engine may be able to improve its CTR predictions by quantifying the content preferences associated with each query across contexts (e.g., time, location, device) based on observed search volume and CTR. Such information could also be valuable for advertisers. Although search-related advertising has reached \$90 billion annually in the United States alone (Statista 2017a), Google (2014) argues that advertisers waste their money on more than half of all Internet ads. Therefore, there is a need for advertisers to optimize the set of queries or keywords on which they bid, the amounts they bid, and the ad copy shown to search engine users. Such optimization can now be performed dynamically over time with the advent of moment marketing companies (such as TVTY) that enable advertisers to instantly launch and optimize search advertising campaigns triggered by offline events such as TV commercials (Liu and Hill 2021). Such exercises rely on being able to predict the CTR for a sponsored search result at a particular point in time, given the query submitted by the user. Especially when there are no or very few data available to predict CTR based on past CTR alone, a model is needed that links users' content preferences to their search queries, the content of the search results to which they are exposed, and their clicking behavior, across contexts. However, to the best of our knowledge, no methodology has yet been developed to address this opportunity.

In this paper, we aim to fill this gap by developing an interpretable model that cannot only identify and quantify users' content preferences across search contexts in a meaningful way, but also predict CTR on the SERP. Our model combines the merits of graphical models, latent factor models, and content analysis

based on probabilistic topic modeling. In essence, our model is a unified probabilistic nonnegative matrix factorization model (Salakhutdinov and Mnih 2008) that can simultaneously leverage and link data on query search volume, CTR on different search results, and the textual content of these results. It estimates a latent vector of content preferences over topics revealed by each query, with preferences allowed to vary systematically across different contexts. It also associates the content of each search query and search result description (title and snippet on the SERP) with a latent vector over the same set of topics, constraining both sets of vectors to be sparse and nonnegative. Therefore, compared with existing applications of natural language processing (NLP) and topic modeling in marketing and recommendation systems (e.g., Gopalan et al. 2014, Tirunillai and Tellis 2014, Liu and Toubia 2018), our proposed approach can simultaneously capture multiple types of information and investigate multiple aspects of behavioral dynamics in a single, robust modeling framework. To facilitate efficient and scalable inference, we develop a full variational Bayesian inference algorithm. Its posterior inference method with data augmentation retains conjugacy and performs efficiently with the sparse data that typically describe user behavior on search engines.

Our proposed modeling framework offers at least two benefits over purely predictive approaches. First, we extract interpretable topics and related parameters that shed light on users' content preferences and how these preferences vary across contexts. Second, our method is able to address the cold-start problem for new links. This is because topic modeling can provide a representation in terms of latent themes discovered from the existing document collection, which allows the algorithm to make meaningful predictions and recommendations for a new link that is not part of the training data, based on its textual description on the SERP. This is not possible with traditional collaborative filtering algorithms, for example, which can only make predictions about items that have been consumed. This feature of our model is particularly important for search engines and advertisers because it is critical for them to improve predictions for new links quickly so they can adjust ranking or search campaign strategies without "wasting" too many valuable impressions testing these new links experimentally. Finally, with our proposed modeling framework and inference algorithm, our model predictions can be applied very efficiently for real-time applications. For example, estimating our model for a moderate corpus (with 1,000 unique search queries) takes around 30 minutes on a personal laptop. Computation time should be lower on more powerful machines such as those



available to search engines and data analytics companies. Once all the model parameters have been estimated, user CTR can be predicted on the fly (in milliseconds) for both warm- and cold-start links.

We validate the proposed model using comprehensive data sets from Microsoft Bing on consumer searches for TV shows (aired in February 2016 and 2017), a large and important category that exhibits intuitive user search dynamics. We find that our content-based search modeling framework has the potential to improve the rankings chosen by Bing. Managerially, we find that our model can provide meaningful interpretation of the content preferences underlying each search query and can identify and quantify whether and how the content preferences revealed by a search query vary systematically based on whether the query is observed before, during, or after a show is aired. Due to the nature of our empirical data sets, in this paper we help the search engine improve its prediction of CTR and hence its ranking algorithm for organic results. Our proposed model can be applied similarly to sponsored search data available to advertisers, for example, to gain insights into which search ads should be targeted to which users, when, and on which devices.

The rest of the paper is organized as follows. We review the relevant literature in Section 2. We then introduce our proposed model framework in Section 3. We describe our empirical data sets in Section 4. This is followed by Section 5, which reports the estimation results and model evaluation. We conclude in Section 6.

## 2. Relevant Literature

Our content-based search model is built upon graphical models,<sup>3</sup> which provide a language to express assumptions about relationships between different variables (e.g., relationships among users' preferences, search queries, subsequent clicks, and a search engine's textual content in our case), as well as topic modeling and Poisson factorization, both of which impose statistical assumptions for inferring these relationships. Our research context lies at the intersection of the information retrieval and marketing literatures. In this section, we briefly highlight how our research contributes to these literatures.

In information retrieval and marketing, most existing research has aggregated users' behavioral signals (i.e., click throughs) over time across all queries and applied it identically for all types of queries (Agichtein et al. 2006). Other studies have built statistical/econometric models of click-through behavior in attempts to measure the impact of different types of keywords and positions of results, but these have either used aggregated information across consumers' search queries or ignored textual information

altogether (e.g., Yang and Ghose 2010, Narayanan and Kalyanam 2015, De los Santos and Koulayev 2017, Abhishek et al. 2018). Another important issue that has been largely ignored in the literature is the context of users' content preferences (e.g., time, location, demographics, and search device). Adomavicius and Tuzhilin (2015) leveraged such contextual information in the context of recommendation systems, and Radinsky et al. (2013) investigated temporal trends and periodicity in online search interests. However, whereas the analysis performed by these authors was specifically designed to focus on temporal variations, our model can accommodate any discrete context that affects preferences, such as the type of device or the geographic location from which the search query was submitted. In addition, these authors did not leverage the textual information in queries and links. In contrast, our approach extracts features from the content of queries and links, which not only enables interpretation of content preferences across contexts but also allows making predictions for new links that were not part of the training data.

There is a stream of studies in marketing that has applied NLP to analyze search data and other user-generated content to understand consumers and generate better marketing intelligence (Archak et al. 2011, Lee and Bradlow 2011, Ghose et al. 2012, Netzer et al. 2012). More recent work has adopted topic modeling to extract topic-level information from content such as online reviews (Tirunillai and Tellis 2014, Büschken and Allenby 2017) and search keywords (Abhishek et al. 2018). However, text-based search behaviors such as typing a search query have not yet been studied extensively. Typing a search query is a first-order search behavior on most search platforms, and queries contain valuable information about users' preferences. Recently, Liu and Toubia (2018) developed a topic model that quantifies how the content in search results relates to the content in search queries, which can be applied to estimate a user's preferences based on a single query. Our work differs from Liu and Toubia (2018) in three main aspects. First, the approach that Liu and Toubia (2018) take relies only on publicly available data (from the Google application programming interface [API]) to link the content in search queries to the content in search results. In contrast, we leverage richer data from Microsoft Bing, which include search volume and CTR over time. Second (and as a result of the first difference), the estimation of content preferences based on the approach of Liu and Toubia (2018) requires specifying assumptions about how users translate their content preferences into search queries. In contrast, our richer data allow us to remain agnostic on the underlying process. Third, methodologically, whereas Liu and Toubia (2018) and other studies on the extant

applications of topic models in the marketing literature have been primarily based on latent Dirichlet allocation (LDA) (Blei et al. 2003), our modeling approach relies on Poisson factorization.

Poisson factorization is an alternative text model to and has been shown to outperform LDA (Canny 2004, Gopalan et al. 2014). It is a form of probabilistic non-negative matrix factorization (Salakhutdinov and Mnih 2008, Cemgil 2009) that replaces typical Gaussian likelihood and real-valued representations with Poisson likelihood and nonnegative representation. Two properties of Poisson factorization make it particularly attractive (Canny 2004). First, by definition, the model can support sparse matrices, which is convenient when dealing with sparse data in real-world problems such as consumer product ratings and movie viewership. Second, by imposing a proper gamma prior distribution, it is possible to build a sophisticated Poisson factorization model that maintains conditional conjugacy, allowing the development of efficient inference algorithms for fitting large-scale data. Poisson factorization has been applied in recommendation system problems for (scientific and news) articles, movies, and music (Gopalan et al. 2013, 2014). To the best of our knowledge, our paper is the first to use Poisson factorization to model user behavior and content simultaneously in the context of search engines.

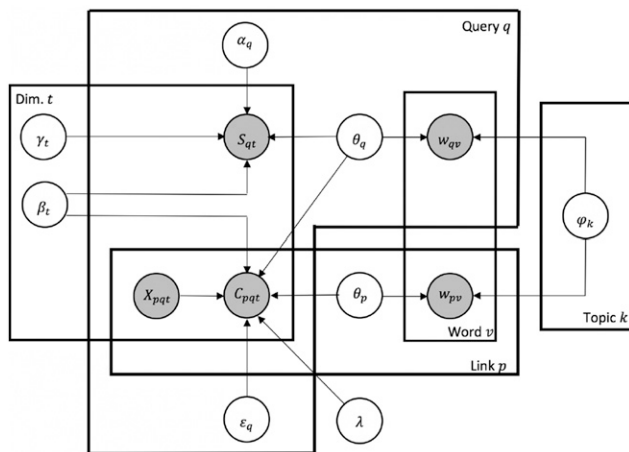
### 3. Content-Based Search Model

In this section, we first describe each component of our model. We then derive its posterior inference and introduce its inference algorithm based on the variational Bayes method.

#### 3.1. Specification

Figure 3 shows the graphical model representation of our approach, which includes five groups of indices:

**Figure 3.** Graphical Representation of the Content-Based Search Model



search queries, indexed by  $q$ ; links, indexed by  $p$ ; words in the vocabulary, indexed by  $w$ ; topics, indexed by  $k$ ; and search contexts, indexed by  $t$ . In our case, context is captured by time: before, during, and after a TV show is aired. This comes without any loss of generalizability. The input to the model (i.e., the variables that are observed) are shaded in Figure 3. The number of occurrences of each word in each query is  $w_{qv}$ ; the number of occurrences of each word in each link description is  $w_{pv}$ ; the search volume for each query in each period is  $S_{qt}$ ; the CTR on each link retrieved as a top result for each query in each period is  $C_{pqt}$ ; and other query-link-time-specific covariates such as position are  $X_{pqt}$ . The remaining variables are unknown model parameters that are defined in the following as we describe the five blocks of the model.

**3.1.1. Topics.** Let  $V$  denote the total number of words in the vocabulary of a particular search category, with these words indexed by  $\{1, 2, \dots, V\}$ . We use a collection of  $K$  unnormalized topics  $\phi_{1:K}$  to describe the latent space of users' content preferences, their queries, and the descriptions of the links presented on the SERP. Each topic  $\phi_k$  is a collection of intensities of the words in the vocabulary, and the prior distribution on each component  $\phi_{vk}$  is gamma:  $\phi_{vk} \sim \text{Gamma}(a_1, a_2)$ , where  $a_1$  and  $a_2$  are the shape and rate parameters for the gamma distribution, respectively.

**3.1.2. Queries.** We consider a collection of  $Q$  target search queries (defined a priori). Each query  $q$  is represented by a vector of length  $V$ ,  $\{w_{q1}, w_{q2}, \dots, w_{qV}\}$ , where  $w_{qv}$  is the number of times the  $v$ th word in the dictionary appears in query  $q$ . To model the generation process of each query's content, each query is associated with a vector of  $K$  latent topic intensities,  $\theta_q$ . The prior distribution on each component  $\theta_{qk}$  is assumed to be gamma:  $\theta_{qk} \sim \text{gamma}(c_1, c_2)$ . Then, the observed word count  $w_{qv}$  is assumed to follow a Poisson distribution whose rate parameter is the inner product of  $q$ 's topic intensities and  $v$ 's topic weights, that is:

$$w_{qv} \sim \text{Poisson}(\theta_q^T \phi_v), \quad (1)$$

for  $q = 1, 2, \dots, Q$  and  $v = 1, 2, \dots, V$ . The form of the factorization  $\theta_q^T \phi_v$  represents the expected number of occurrences of word  $v$  in query  $q$ . The choice of the gamma distribution for the prior and the Poisson distribution for the likelihood function offers several benefits (Gopalan et al. 2013). First, the gamma distribution induces sparsity on the topic weights and intensities. Second, the Poisson distribution is well-suited for sparse data like ours. Third, these two distributions are conjugate, which provides significant computational benefits.

**3.1.3. Links.** We consider the collection of  $P$  different links retrieved by the set of  $Q$  queries. We focus on the information available to search engine users when deciding whether to click on each link. Accordingly, for each link, we focus on the text of its description (including the title) on the SERP, rather than the content of the actual web page to which the link points. Each link description  $p$  is represented by a vector  $w_p$  of length  $V$ ,  $\{w_{p1}, w_{p2}, \dots, w_{pV}\}$ , where  $w_{pv}$  is the number of times the  $v$ th word in the dictionary appears in the description of link  $p$  shown to users on the SERP.<sup>4</sup> The generation process of the link description is defined similarly as that for queries. We define each link  $p$ 's topic intensities as  $\theta_p$ :  $\theta_{pk} \sim \text{gamma}(b_1, b_2)$ . The observed word count is assumed to satisfy:

$$w_{pv} \sim \text{Poisson}(\theta_p^T \phi_v), \quad (2)$$

for  $p = 1, 2, \dots, P$  and  $v = 1, 2, \dots, V$ . The form of the factorization  $\theta_p^T \phi_v$  represents the expected number of occurrences of word  $v$  in link  $p$ .

**3.1.4. Search Volume.** For each query  $q$  issued at time  $t$ , we observe two types of user search behaviors, search volume and CTR on the top search results. We model search volume (the number of times each query  $q$  is issued by users at time  $t$ ) as a Poisson distribution, whose rate parameter couples users' time-specific content preferences,  $\beta_t$  and  $\gamma_t$ , with the query's topic intensities,  $\theta_q$ :

$$S_{qt} \sim \text{Poisson}(\alpha_q + \theta_q^T (\beta_t + \gamma_t)). \quad (3)$$

In this equation,  $\alpha_q$  captures the baseline popularity or search volume for query  $q$ ,  $\theta_q$  contains the topic intensities of the query introduced in Equation (1), and  $\beta_t$  and  $\gamma_t$  capture temporal variations in users' content preferences. The distinction between these two parameters is discussed in more detail following the introduction of the next equation. For the same reason mentioned earlier when introducing the prior distribution for the query-related parameters, the parameters  $\alpha_{qk}$ ,  $\beta_{tk}$ , and  $\gamma_{tk}$  also have a gamma prior distribution:  $\text{Gamma}(e_1, e_2)$ .<sup>5</sup>

**3.1.5. Click-Through Rates.** When  $S_{qt} > 0$ , the search engine will retrieve results for query  $q$ . We let  $I_{pqt}$  denote the observed number of impressions of link  $p$  on the first SERP for query  $q$  during period  $t$ . We treat multiple clicks on the same result by the same user following a query as only one click, so users' total number of clicks  $\text{Click}_{pqt}$  on link  $p$  at time  $t$  is bounded by  $I_{pqt}$ . We observe  $\text{Click}_{pqt}$  only if  $S_{qt} > 0$  and  $I_{pqt} > 0$ . Let  $L_{qt}$  denote all the top links for query  $q$  observed at time  $t$ . For each link  $p \in L_{qt}$ , we model its CTR using a

Poisson distribution. As the Poisson distribution captures only discrete counts, we model the rounded number of clicks that link  $p$  receives per 100 impressions (i.e., the integer part of  $\text{Click}_{pqt}/I_{pqt} \times 100$ , denoted as  $C_{pqt}$ ). We model CTR as being related to both query-specific and time-specific user content preferences as follows:

$$C_{pqt} \sim \text{Poisson}(\theta_p^T (\theta_q + \varepsilon_q + \beta_t) + \lambda_{X_{pqt}^{\text{pos}}}), \quad (4)$$

where  $\varepsilon_q$  is part of the user content preferences underlying query  $q$ , which we discuss in more detail in the following paragraph; the positive scale parameter  $\lambda_j$  captures the fixed effect of position  $j$  on the SERP; and  $X_{pqt}^{\text{pos}}$  is an integer indicating the (rounded) observed average position of link  $p$  for query  $q$  at time  $t$  on the SERP.<sup>6</sup> We note that our model assumes, for tractability, that the CTR of each link on a SERP is independent of the other links on the page, which is a limitation. Our results provide a lower bound of the performance that could be achieved by extensions of our model that relax this assumption, which we leave to future research.

The user content preferences underlying query  $q$  at time  $t$  are specified as  $\theta_q + \varepsilon_q + \beta_t$ . The presence of the topic intensities of query  $\theta_q$  reflects the fact that user content preferences underlying query  $q$  should be at least partly captured by the actual content of the query. The presence of the vector of offset parameters  $\varepsilon_q$  is motivated by two factors. First, it scales the parameter  $\theta_q$  to fit the CTR data (as  $\theta_q$ , inferred mainly through the text in query  $q$ , tends to have very small magnitude). Second, it allows content preferences across topics to deviate from the textual information in the query. Such specification is consistent with Liu and Toubia (2018, 2020), who suggested that queries may not be a direct representation of what users are searching for. Hence,  $\theta_q + \varepsilon_q$  captures the query-specific, time-invariant preferences underlying query  $q$ . We assume that  $\varepsilon_q$  also has a gamma prior distribution:  $\text{Gamma}(e_1, e_2)$ .

Finally, we discuss the difference between parameters  $\gamma_t$  and  $\beta_t$ , which capture different types of temporal variations in content preferences. Parameter  $\gamma_t$  does not appear in the CTR equation (Equation (4)). Hence,  $\gamma_t$  captures temporal variations in users' content preferences that are directly related only to changes in search volume. Holding the query constant,  $\gamma_t$  does not influence CTR for the links retrieved by that query. If a topic has a high weight  $\gamma_t$  in a particular period  $t$ , Equation (3) implies that a query that features this topic will tend to have a higher search volume during the same period. However, the fact that a particular link description features this topic does not influence the CTR for the link, conditional on the query. The parameter  $\gamma_t$  may, however, influence

the unconditional CTR on link descriptions that feature that topic at time  $t$ , but only through variations in search volume. The inclusion of  $\gamma_t$  is motivated by variations in the data illustrated by Figure 1; in other words, this set of parameters captures variations in content preferences that are captured by variations in search volume.

In contrast, the set of parameters  $\beta_t$  appears in both Equations (3) and (4) and thus these parameters capture temporal variations in users' content preferences that are directly related to variations in both search volume and CTR. That is,  $\beta_t$  allows the CTR on links to vary over time, even when holding the query constant. The presence of these parameters in the model is motivated by the model-free evidence illustrated by Figure 2, which suggests that variations in search volume are insufficient to provide a full picture of the variations in content preferences, and that CTR may vary over time for a (query, link) pair. In our empirical study reported in Section 5, we provide additional evidence to support our model specification.

### 3.2. Remarks

In this paper, we apply this model on search logs from the Bing search engine for organic results. In doing so, we aim to help the search engine improve its prediction of CTR and hence its ranking algorithm. However, we note that our model can also be applied to sponsored search data available to both search engines and advertisers. In particular, advertisers get reports from search engines on query impressions for their campaigns, the average positions of their search ads, and CTRs on different ads across contexts. Once all posteriors of the model parameters are calibrated (details are given in Section 3.3), the first term,  $\theta_p^T(\theta_q + \varepsilon_q + \beta_t)$ , in Equation (4) predicts the relative click-through tendency on ad  $p$  given search query  $q$  at  $t$  by capturing the fit between the ad copy and the estimated content preferences revealed by the query. This prediction can help advertisers design effective campaigns so that more preferred web pages are promoted, with descriptions that resonate better with users.

We have two additional notes regarding our proposed framework. First, our model can predict click-through tendency for a (query, time, link) combination even for cold-start links (i.e., links that were never observed in the training data). Such prediction requires the parameters  $\theta_p$ ,  $\theta_q$ ,  $\varepsilon_q$ , and  $\beta_t$ . These last three sets of parameters are estimated based on the training data, whereas the topic intensities of the link description  $\theta_p$  may be estimated based on the content of the link description, using the model parameters estimated from the training data. Second, our model is flexible in the sense that depending on the nature of available data and the research objective, one can add

or remove some components while still maintaining model conjugacy (which is described in Section 3.3). For example, one can have multiple indices  $t$  such as time and search device, or one can remove the search volume equation.

### 3.3. Posterior Inference and Inference Algorithm

The central computational problem is posterior inference, which reverses the model's generative process to learn the unknown parameters. In our application, the input of the model includes the content of queries and link descriptions on a search engine, the search volume for each query over time, the CTR, and the average position for each (query, time, link) combination. Given  $\{w_q, w_p, S_{qt}, C_{pqt}, X_{pqt}^{pos}\}$ , our goal is to infer the model parameters, including  $\{\varphi_k\}$ ,  $\{\varepsilon_q\}$ ,  $\{\gamma_t\}$ ,  $\{\beta_t\}$ ,  $\{\theta_q\}$ ,  $\{\theta_p\}$ ,  $\{\alpha_q\}$ , and  $\{\lambda_j\}$ .

To facilitate inference, we first augment the model with auxiliary variables in Equations (1) to (4), which makes it conditionally conjugate. In Appendix A, we show how these auxiliary variables fit into the posterior inference and derive the full conditionals for all the model parameters. Preserving conjugacy provides a closed-form understanding of how the algorithm places different weights across observed data in learning model parameters. More importantly, it also leads to significant computational benefits, even with variational Bayes inference. When deriving some of these posteriors, we also discuss the underlying intuition and some attractive features resulting from our (conjugate) model specification. For example, our inference procedure iterates only over nonzero count observations, which leads to significant computational efficiency for large-scale data. In addition, our model inference naturally addresses the sparsity of the textual information in search queries by augmenting them with CTR data in a semisupervised fashion. This can ultimately enhance the interpretability/learning of the topic distributions and the underlying consumer content preferences behind different search queries. Readers should refer to Appendix A for additional technical details.

One could simply run a Gibbs sampler to iterate over the posterior of all the model parameters given in Appendix A. However, given the large-scale nature of real-world online search data in this problem, the computational time required for Markov chain Monte Carlo (MCMC) methods would be prohibitive (Salakhutdinov and Mnih 2008). To address scalability, we adopt a variational inference algorithm, which is a deterministic optimization-based strategy for approximating posterior distributions in complex and large-scale probabilistic models (Jordan et al. 1999). The basic idea of variational inference is that it posits a family of distributions over the hidden variables, indexed by free variational parameters,



and then finds the member of this family that is closest in Kullback-Liebert divergence to the true posterior. The variational inference algorithm typically converges to its final approximation much faster than MCMC methods, and convergence can be assessed more easily than with MCMC methods. In this research, we derive a coordinate ascent variational inference (CAVI) algorithm, which is based on the mean-field family in which all the latent variables are assumed to be independent and each is governed by its own distribution (Jordan et al. 1999). Appendix B describes the details of this algorithm. Note that because of the conditional independence of the variables shown in the posterior inference, the data can be distributed across many machines, and the posterior updates for each type of variable can be implemented in parallel to improve convergence speed.

## 4. Data

### 4.1. Data Collection and Processing

Our empirical study is conducted with data from search logs for TV shows from the Bing search engine in the U.S. market. There are several reasons why we use this data set for our empirical study. First, TV show search is an important category, as millions of users search for TV shows on a daily basis (Statista 2017b). Second, TV show searches exhibit intuitive dynamics that we can use to evaluate the face validity of the model output. In particular, we can expect users' interest in a TV show to vary depending on whether the show has aired (Fossen and Schweidel 2017).

We study 14 TV shows from various genres (including sports, variety, reality, drama, and comedy) that were aired in February 2016 and were popular according to Comcast's viewership rankings across all TV shows during that month. Table 1 provides the

show names and broadcast networks. For each show, we collect related searches whose queries contain either the name of the show or names of major cast members obtained from IMDB. We focus on searches that were issued over three separate time windows for a given episode: the 24 hours before the show was aired, during the show, and the 24 hours after the show was aired. Hence, we study the dynamics of user preferences across  $T = 3$  periods in this application. We collect all searches across multiple episodes aired in February 2016 for 11 weekly TV shows. For three annual events (i.e., the Super Bowl, Grammy Awards, and Oscars), we collect all searches for two consecutive years (2016 and 2017). We treat each episode as a separate set of observations, and pool information across episodes to train the model parameters for each show separately.

We focus on the queries that were issued by users a minimum number of times within each time window for each episode (the cutoff value is five for weekly TV shows and 20 for annual TV events).<sup>7</sup> For each selected search query observed in each time window, we collect all the organic links on the first SERP of Bing. To obtain reliable measures of CTR, for each query  $q$  observed at time  $t$ , we include only links that were presented on the first SERP in at least 60% of the impressions. For each (query, time, link) combination, we compute the average position of the link across impressions when training the model (however, all of our model validation is done at the level of individual impressions). When computing the CTR of each (query, time, link) combination, we incorporate the following empirical decisions. First, when multiple links are clicked after a search by the same user, we consider all the clicks. Second, if the same link is clicked multiple times after a search by the same user,

**Table 1.** Descriptive Statistics of the Corpora

ID. Show name (network)	Episodes	Vocabulary size	Queries	Words per query	Links	Words per link	Query-time combinations	Links per query-time combination
1. Super Bowl (CBS)	2	3,552	5,631	3.59 (1.10)	8,634	23.69 (8.50)	9,026	6.64 (2.68)
2. Oscars (ABC)	2	1,840	2,231	3.19 (0.95)	3,520	21.60 (8.00)	3,442	6.67 (2.69)
3. Grammy (CBS)	2	1,069	1,502	3.29 (1.06)	1,996	21.04 (7.74)	2,215	6.12 (2.45)
4. Shark Tank (ABC)	3	423	122	3.64 (0.86)	568	14.55 (4.37)	250	7.88 (2.96)
5. The Bachelor (ABC)	3	1,074	785	2.91 (0.88)	1,979	15.44 (3.92)	1,211	7.55 (2.73)
6. Saturday Night Live (NBC)	2	1,342	650	3.69 (1.41)	1,983	17.11 (4.50)	848	6.77 (2.40)
7. Survivor: Kaôh Rong (CBS)	2	1,070	407	2.89 (1.10)	1,548	14.74 (4.24)	807	7.25 (2.73)
8. The X-Files (Fox)	4	892	523	2.60 (0.93)	1,549	14.60 (4.11)	1,195	7.11 (2.77)
9. The Blacklist (NBC)	4	409	201	2.86 (1.13)	701	13.53 (4.26)	616	7.30 (2.56)
10. The Walking Dead (AMC)	3	1,750	2,155	3.96 (0.97)	4,338	17.75 (3.96)	4,805	6.84 (2.34)
11. Criminal Minds (CBS)	2	338	130	3.78 (0.96)	526	15.25 (4.14)	257	7.85 (2.76)
12. Law & Order (NBC)	4	829	276	3.75 (1.48)	1,229	16.58 (5.22)	587	7.04 (1.98)
13. The Big Bang Theory (CBS)	2	849	317	3.85 (1.30)	1,240	16.06 (4.45)	608	6.83 (2.36)
14. Modern Family (ABC)	3	263	71	3.08 (0.86)	386	14.08 (4.72)	222	7.81 (2.86)

*Notes.* We report the average with the standard deviation in parentheses. The length of a query or link description is calculated based on the vocabulary.

we count this as one click. Finally, we combine all the selected queries, their search results, and link descriptions for the same show as one corpus. Following the bag-of-words approach, we treat each query or link description as an unordered set of words. We process the text in each corpus based on standard practice in text mining. We remove any delimiting characters used to separate words; we eliminate punctuation and a standard list of English stop words; and we do not perform stemming. We form the vocabulary for each corpus using words that appear at least five times across all the documents.

We now report some descriptive statistics of the corpora, subject to confidentiality constraints that prevent us from reporting raw search volume or CTR data. First, Table 1 reports the size of the corpora, including the vocabulary size, number of unique queries/links, number of words per query/link description, number of unique (query, time) combinations, and number of links per (query, time) combination. In addition, we illustrate the overall sparsity in CTR at the query-link-time level, which is the unit of analysis in the model, for annual TV events and weekly shows separately. For weekly shows, we find that 48.1% of observations have a CTR of 0 (with a standard deviation of 5.5% across TV shows). For annual events, about 25.5% of observations have a CTR of 0 (with a standard deviation of 2.1%). These observations confirm the sparsity in CTR. This further suggests the challenges in extracting meaningful information from consumer search behavior and hence in predicting CTR. Finally, online Appendix C presents additional descriptive statistics related to the variation across searches and position effect on CTR.

#### 4.2. Additional Data for Assessing Search Result Variations

Although Bing did not strategically adjust search results around airing time during our main study period, we want to understand more generally the systematic variations in search engines' rankings. Hence, we further explore the systematic variations in the appearance as well as position of search results before, during, and after TV events. For that, we collect additional data from both Google and Bing for TV shows. As these additional data were collected at a different time than our main data, we acknowledge the possibility that their results may not hold in our main study period. All other analyses reported in the paper are based on our main data described in Section 4.1. We want to emphasize that our model takes the search engine results as given and is agnostic as to whether variations exist in search results around the airing time of shows as well as the source of this variation. In Section 5.4, we compare the ranking of

results suggested by our model to the actual ranking on Bing.

**4.2.1. Data Collection.** We compiled a list of 50 top weekly TV shows and 29 sports events (i.e., NBA and NFL games) scheduled to be aired during November 2018. For each weekly TV show, we formed one search query that contained only the show name; for each sports event, we formed one search query that contained the two teams playing against each other (e.g., "Seattle Seahawks VS Green Bay Packers"). We wrote a Python script that automatically collected the top 10 organic search results for these 79 search queries on Bing and Google, every 30 minutes continuously (i.e., 48 times per day for each search engine). The script was run for a month, from November 4, 2018 at 7:15 a.m. to December 5, 2018 at 7:15 a.m. During this period, there were a total of 212 TV events: 183 unique episodes from the 50 weekly TV shows plus 29 sports events, for which we also collected the actual airing (start and end) times. Although the data acquired from the APIs do not reflect customization at the individual user level, they do reflect the baseline SERP from a given search engine for a given query at a particular point in time.

**4.2.2. Identification Strategy.** Our objective is to measure systematic variations in the results that are shown in response to a search query for a TV show around the show airing time. For each event  $e$  on a given search engine and each link  $l$  that appeared at least once for event  $e$ , we let  $I_r^{e,l}$  indicate whether link  $l$  appeared on the top SERP in the API request  $r$ ; if  $I_r^{e,l} = 1$ , we let  $Pos_r^{e,l}$  denote the position of link  $l$  in API request  $r$ . We are interested in the amount of variation in  $I_r^{e,l}$  and  $Pos_r^{e,l}$  across three windows: (1) the  $X$  hours before the start of show airing, (2) during show airing, (3) the  $X$  hours after the end of show airing. If the search results are adjusted around the airing of the event, the variations between these three windows should be increased due to the airing of the event. We consider  $X \in \{12, 24, 36\}$  to evaluate the robustness of our results.

For each  $(e, l)$ , we compute five statistics to capture variations in  $I_r^{e,l}$  and  $Pos_r^{e,l}$ :

- $Y_{1a}^{e,l}$ : a binary variable equal to 1 if link  $l$  appeared in only one of the three windows.
- $Y_{1b}^{e,l}$  (defined only if  $Y_{1a}^{e,l} = 0$ ): a binary variable equal to 1 if the total variance of  $I_r^{e,l}$  over the API requests corresponding to the three windows is positive, that is,  $S_{\text{Total}}^{e,l} = \sum_{w,r \in O_w^e} (I_r^{e,l} - \bar{I}^{e,l})^2 > 0$ , where  $O_w^e$  is the set of API requests within time window  $w$  for event  $e$ , and  $\bar{I}^{e,l}$  is the average of  $I_r^{e,l}$  over the three windows.
- $Y_{1c}^{e,l}$  (defined only if  $Y_{1b}^{e,l} = 1$ ): the ratio of the between-window variation in  $I_r^{e,l}$  to the total variation,

i.e.,  $Y_{1c}^{e,l} = \sum_w |O_w^e| (\bar{I}_w^{e,l} - \bar{I}^{e,l})^2 / S_{\text{Total}}^{e,l}$ , where  $\bar{I}_w^{e,l}$  denotes the average of  $I_r^{e,l}$  in window  $w$ .

- $Y_{1d}^{e,l}$  (defined only if  $Y_{1a}^{e,l} = 0$ ): an indicator variable equal to 1 if the total variance of  $Pos_r^{e,l}$  over the API requests corresponding to the three windows is positive. The formula for the total variance is the same as for  $Y_{1b}^{e,l}$ , replacing  $I_r^{e,l}$  with  $Pos_r^{e,l}$  and only counting API requests for which  $I_r^{e,l} = 1$ .

- $Y_{1e}^{e,l}$  (defined only if  $Y_{1d}^{e,l} = 1$ ): the ratio of the between-window variation in  $Pos_r^{e,l}$  to the total variation. The formula is the same as for  $Y_{1c}^{e,l}$ , replacing  $I_r^{e,l}$  with  $Pos_r^{e,l}$ .

If results change around show airing,  $Y_{1a}^{e,l}$  will more likely to be equal to 1. If  $Y_{1b}^{e,l} = 1$  and/or  $Y_{1d}^{e,l} = 1$ , it means that there is some variation in the set of search results and their positions. A large value of  $Y_{1c}^{e,l}$  and/or  $Y_{1e}^{e,l}$  indicates systematic variations across windows, relative to the overall amount of variation in search results.

Although some variations in search results are bound to occur naturally over searches, our focus is specifically on variations that happen due to the airing of the event. That is, ideally, for each  $(e, l)$  pair we would like to compare the presented statistics from the observed data to those from a counterfactual set of observations over the same time period but without the airing of the event. Hence, for each statistic, we are interested in the following treatment effect:  $Y_1^{e,l} - Y_0^{e,l}$ , where the subscript 0 refers to a counterfactual scenario in which the TV event would not be scheduled during the time window in which it was scheduled. Because we only observe  $Y_1^{e,l}$ , we use two pseudo-control conditions to approximate  $Y_0^{e,l}$ . They have the same time duration as the treated condition, but they are shifted by three days either backward (denoted as “control pre”) or forward (denoted as “control post”). Hence, the control conditions are similar to the treated one for each TV event, with the important distinction that only the treated period includes the actual airing of the event (which also required having the control periods cover different days of the week).

As an example, we illustrate the operation for a TV event aired on November 10 at 8:00 p.m. for 30 minutes when  $X = 24$ . In this case, the treated observations cover the following three time periods: November 9, 8:00 p.m.–November 10, 7:59 p.m. (24 hours before); November 10, 8:00 p.m.–November 10, 8:30 p.m. (during); and November 10, 8:31 p.m.–November 11, 8:30 p.m. (24 hours after). The first control window (control pre) would be similar to the treatment window, but shifted by three days into the past: November 6, 8:00 p.m.–November 7, 7:59 p.m.; November 7, 8:00 p.m.–November 7, 8:30 p.m.; and November 7, 8:31 p.m.–November 8, 8:30 p.m. The

second control window (control post) would be shifted by three days in the other direction: November 12, 8:00 p.m.–November 13, 7:59 p.m.; November 13, 8:00 p.m.–November 13, 8:30 p.m.; and November 13, 8:31 p.m.–November 14, 8:30 p.m. Note that for all TV events and values of  $X$ , the control windows never include any airing of the show, as our events were either one-time events or weekly TV shows.

**4.2.3 Results.** We compute the five statistics in the treated and control conditions, and take the average across links and across TV events of the same type (TV show versus sports event), for each search engine and for each value of  $X$ . The details are presented in online Appendix D. Here we provide only a brief summary of the comparisons between the treated and control conditions. We find some evidence that the identity and position of search results vary systematically around the airing time of TV shows and sports events on Google and Bing, with the exception of TV shows on Bing. Without access to internal data from Google, we cannot determine whether these variations are driven by variations in content, search volume, and CTRs that lead to automatic and mechanical updating of the results or by a strategic and conscious effort by the search engine. In the case of Bing, our discussions with the company suggest that the search engine did not strategically update the set of results and their positions around TV airing times in our main study period. Again, these findings are not directly relevant to the results reported in the following section, in which we compare the ranking of results suggested by our model to the actual ranking on Bing, without making any assumption on the extent or source of variations in search results around the airing time of shows.

## 5. Empirical Results

In this section, we first describe the model estimation procedure using our data, and then report sample model output that may provide interesting and important insights for managers. Lastly, we evaluate the performance of our model in CTR prediction and ranking recommendation.

### 5.1. Estimation

We split the corpus for each TV show into in-sample and out-sample data sets (recall that we estimate the model separately for each TV show). If a TV show has  $N$  episodes in our data set, the in-sample data set contains all the searches from the first  $N - 1$  episodes and the out-sample data set includes all the searches from the last episode. This allows us to simulate a situation in which the search engine would update its predicted CTR and optimize the position of results for



searches around the airing time of a new episode for each  $(q, t)$  combination, based on the searches from previous episodes. We further split the in-sample data set into a training data set, which is used for estimating the model parameters, and a tuning data set, which is used for determining the optimal choice of the number of topics  $K$ . We form the tuning data set from 10% of randomly selected (query, time, link) combinations, leaving the remaining 90% as the training data set. In the out-sample data set, about 25% (70%) of queries for annual events (TV shows) also appear in the in-sample data set (i.e., over the past episodes). These warm-start queries account for about 50% (90%) of searches in the out-sample data set for annual events (TV shows). Note that our proposed model can only make predictions for warm-start queries as these predictions rely on query-specific parameters trained using in-sample observations. However, our model is able to handle the cold-start link problem. In the case of a cold-start link, when predicting CTR out of sample, we estimate the topic intensities of the new link by running a Poisson factorization, using the textual description of the link and the topics extracted from the training stage.

Following Gopalan et al. (2013), we set each gamma shape and rate hyper-parameter at 0.3 for topics, topic distribution, preferences, and baseline search volume. For the scale parameters  $\lambda$ , we set the gamma shape and rate hyper-parameter at 0.1 and 1, respectively. We initialize topics  $\phi$  and topic distribution  $\theta$  for queries and links using LDA. We search for  $K$  in [5, 20]. For each given candidate  $K$ , we estimate our proposed model using the training data set, and we evaluate the model performance on the tuning data set. We define the optimal  $K$  as the one that yields the highest accuracy in predicting the CTR in the tuning observations. The specific metric we use is the mean absolute error between actual and predicted CTR. Our selected  $K$  varies between 6 and 14 for various TV shows. More generally, firms that apply our model can set  $K$  to be smaller or larger, depending on the trade-off between clear and intuitive interpretation on the one hand and prediction accuracy on the other.

## 5.2. Substantive Results

As an example, in this section we report the substantive results from the Super Bowl corpus, which show very good face validity. These observations confirm that our model may provide reliable insights for other search domains. Importantly, as mentioned earlier, these insights would not be possible with a purely item-based modeling approach to ranking, as such an approach would not model the content of queries and search results.

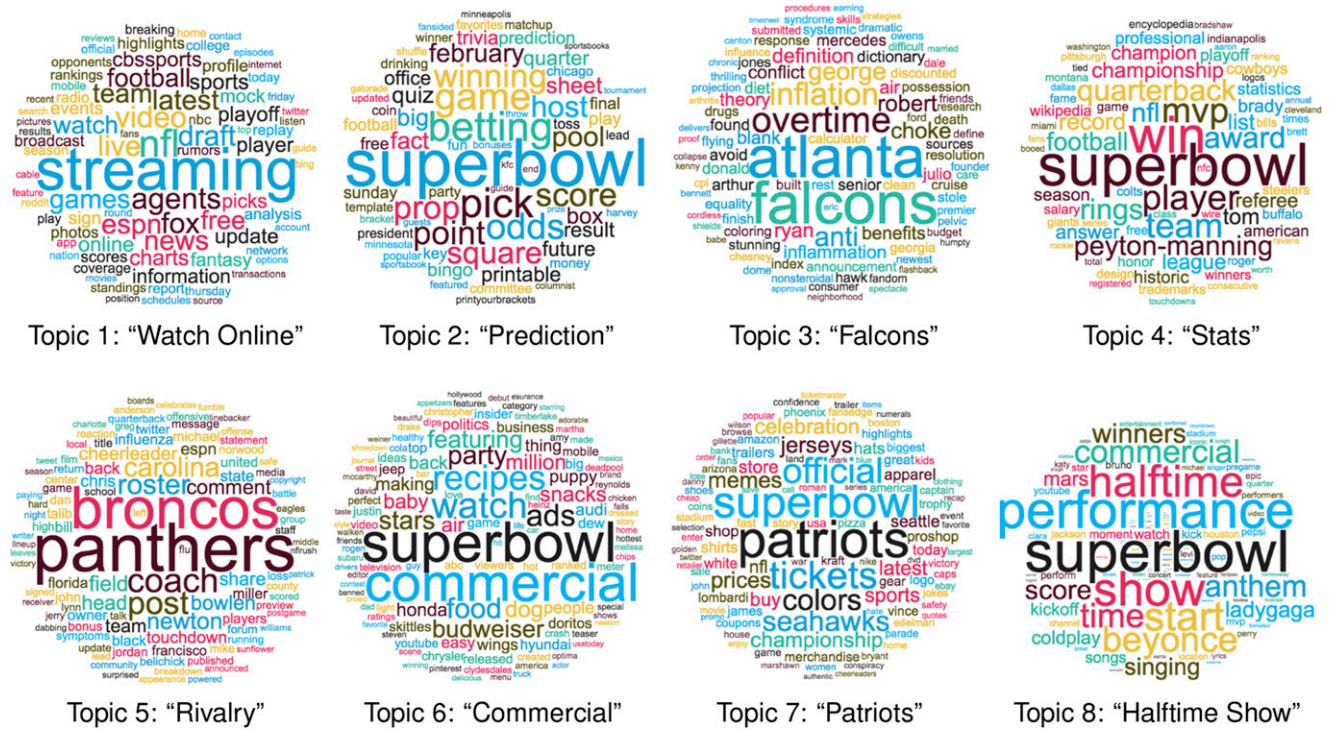
**5.2.1 Topics.** The first output is the topics that are extracted from a corpus. To ease interpretation, we compute the relevance of each word in each topic (Bischof and Airolidi 2012, Sievert and Shirley 2014).<sup>8</sup> Then, we simulate content for each topic using the exponential of the computed relevance measure. That is, we generate sets of words for each topic, where the occurrence probability of each word is proportional to the exponential of its relevance. We use word clouds to visualize the simulated sets of words, where words with larger font size have higher relevance.

Figure 4 present all eight topics from the Super Bowl. In plotting these word clouds, we have removed words that are very common or not highly interpretable (e.g., non-English terms, verbs, adjectives) for better visualization. The sets of the most relevant words are quite different across topics, and the topics revealed through these words are coherent and meaningful. These extracted topics also confirm our expectations and knowledge about what content users may be interested in regarding the Super Bowl. We label these eight topics “Watch Online,” “Prediction,” “Falcons,” “Stats,” “Rivalry,” “Commercial,” “Patriots,” and “Halftime Show.” In online Appendix E, we report eight topics from the popular TV show *The Walking Dead*. Overall, we find that most topics are show-specific, but there are a few topics that appear consistently across most TV shows. These are mainly about show time and watching online, spoilers and recaps, and episodes and seasons.

**5.2.2. Time-Specific Content Preferences.** We analyze the estimated time-specific variations in content preferences directly reflected in search volume only (i.e., those that do not influence CTR, controlling for the query),  $\gamma_t$ , and variations that are directly reflected in both search volume and CTR,  $\beta_t$ . As an example, Table 2 reports these estimated topic weights for the Super Bowl. It is clear that there are major differences between these two types of time-specific content preferences. For instance, there are more topics with large weights on  $\gamma_t$  than on  $\beta_t$ . That is, variations in content preferences seem to manifest mostly through variations in search volume. For instance, users are more likely to search for information on how to watch a game online and on predictions (topics 1 and 2) before the game; teams playing in the game (topics 3 and 5) during the game; and game results, commercials, and halftime shows (topics 4, 6, and 8) after the game. Nevertheless,  $\beta_t$  is large for some topic-time combinations, meaning that some variations in preferences may only be identified from variations in CTR. For example, holding the query constant, users are relatively more likely to click on links about prediction (topic 2) before the game (recall that only  $\beta_t$  and not  $\gamma_t$  appear in Equation (4), which models CTR).



**Figure 4.** (Color online) Word Clouds of Eight Topics from Super Bowl



*Notes.* Each word cloud contains the simulated content for one of the eight topics from the Super Bowl. The size of each word is proportional to the exponential of its relevance.

We also look at the extracted time-specific content preferences for the 13 other TV shows. As another example, in online Appendix E, we present the estimated time-specific content preferences from *The Walking Dead*. Although the results are show-specific, a few observations are relatively consistent across TV shows. For example, people are more likely to search for specific episodes/seasons before or after a show airs. People are also more likely to search for recaps before and during a show. These general patterns have good face validity.

**5.2.3. Query-Specific Content Preferences.** We can understand the content preferences underlying each

query  $q$  by looking at the topic intensities of query  $q$ ,  $\theta_q$ , which are learned through the text of the query, its search volume, and CTR ( $\theta_q$  appears in Equations (1), (3), and (4)), as well as by examining the offset parameters  $\varepsilon_q$ , which are mainly learned through CTR, as they appear only in Equation (4). First, we find that  $\varepsilon_q$  is often significant, confirming the need for this additional offset parameter. For example, for the Super Bowl, across all queries, the average of the magnitude of the sum of  $\varepsilon_q$  (across topics) is 13.76, with a standard deviation of 9.65. In comparison, the average magnitude of the sum of  $\theta_q$  is 0.29, with a standard deviation of 2.27. Second, we find that  $\theta_q$  and  $\varepsilon_q$  have different topic intensities, confirming that

**Table 2.** User Time-Specific Preferences for the Super Bowl

Topic	$\beta_t + \gamma_t$			$\beta_t$			$\gamma_t$		
	Before	During	After	Before	During	After	Before	During	After
1. Watch online	1,733.5	0.0023	0.0022	$\approx 0$	$\approx 0$	$\approx 0$	1,733.5	0.0023	0.0022
2. Prediction	363.1	0.0034	0.0034	1.3088	0.0001	0.0001	361.8	0.0033	0.0033
3. Falcons	0.0022	1,698.6	0.0021	$\approx 0$	$\approx 0$	$\approx 0$	0.0021	1,698.6	0.0021
4. Stats	0.0119	0.0122	2,407.9	$\approx 0$	$\approx 0$	$\approx 0$	0.0118	0.0122	2,407.9
5. Rivalry	3,329.1	1,542.3	0.0177	0.0001	0.0001	0.0001	3,329.1	1,542.3	0.0176
6. Commercial	0.0387	0.0380	8,216.6	0.0001	0.0001	0.0001	0.0386	0.0379	8,216.6
7. Patriots	0.0005	0.0005	945.0	$\approx 0$	$\approx 0$	$\approx 0$	0.0004	0.0004	945.0
8. Halftime Show	4,017.8	4,048.3	7,893.8	2.1141	0.8266	0.6429	4,015.7	4,047.4	7,893.2

*Notes.* The parameter  $\gamma_t$  captures temporal differences in content preferences that are reflected directly by changes in search volume only; and  $\beta_t$  captures temporal differences in content preferences that are reflected directly in both search volume and in CTR. We use  $\approx 0$  to denote a topic weight that is smaller than 0.0001.

it is important to have both parameters to better infer and understand content preferences at the query level. In particular, the average cosine similarity between  $\theta_q$  and  $\varepsilon_q$  across all the Super Bowl queries is 0.35, with a standard deviation of 0.32.<sup>9</sup>

To further illustrate these differences, Table 3 presents three sample queries and their normalized  $\theta_q$  and  $\theta_q + \varepsilon_q$  (so that the summation of the intensities across topics is 1 for each query). For the first query, “live super bowl stream,”  $\theta_q$  unsurprisingly has dominant intensities on topic “Watch Online.” Although  $\theta_q + \varepsilon_q$  also has the highest weight on the same topic, it also has a higher weight on the topic “Stats.” For the query “facts about the super bowl,” although  $\theta_q$  has a dominant weight on the “Prediction” topic,  $\theta_q + \varepsilon_q$  tends to spread the weights more evenly across topics. Finally, for the query “NFL MVPs,” for which the underlying user search intent is relatively more straightforward, we see that both  $\theta_q$  and  $\theta_q + \varepsilon_q$  have dominant weights on the “Stats” topic, although  $\theta_q + \varepsilon_q$  again tends to spread the weights more evenly across topics. Overall, the extracted user content preferences underlying these search queries confirm our expectations.

**5.2.4. Query-Time Specific Content Preferences.** By combining  $\theta_q + \varepsilon_q$  with  $\beta_t$ , we obtain an estimate of the content preferences associated with query  $q$  at time  $t$ . As an example, we pick the query “super bowl,” which was used to motivate our research in the introduction, and report its estimated  $\theta_q + \varepsilon_q + \beta_t$  in Table 4. We see that within each time window, several topics have large intensities, indicating that users tend to click on very diverse content when searching for “super bowl.” In addition, we see some meaningful shifts in the importance weights across time for the same topic. For instance, content about statistics and the halftime show (topics 4 and 8) is more important before the game starts; content about watching online (topic 1) becomes the most important relative to other topics during the game; and content

about commercials (topic 6) becomes much more important only after the game. The overall pattern of these results is consistent with the click-through data reported in Figure 2. This again suggests that our model can automatically extract meaningful and interpretable insights from web search behavior data.

### 5.3. Predicting Clicks at the Impression Level

We now evaluate our model’s predictions at the level of individual search impressions. In our data set, about 70% of search impressions generate no click on any of the top organic search results whereas the remaining 30% generate at least one click. We test our model’s ability to predict whether a search impression will result in at least one click. To do this, for each search impression we compute the average predicted CTR across search results and compare this average for search impressions for which there was at least one click, denoted as  $CTR^1$ , with the average across search impressions that resulted in no click, denoted as  $CTR^0$ . Table 5 reports the results for weekly shows and annual events separately, for the in-sample and out-sample data sets. Based on two-sample t-tests, we find that  $CTR^1$  is always significantly larger than  $CTR^0$  across all show types and data sets ( $p$ -value  $< 0.001$ ). That is, the average predicted CTR across search results is a good predictor of whether a search engine user will click on at least one of the search results.

### 5.4. Improving the Search Engine’s Ranking

For each search impression, our model can also be used to order links based on predicted CTR, giving rise to a ranking recommendation that may differ from Bing’s ranking. In this section, we aim to determine whether and under what circumstances our content-based approach can help Bing improve its ranking. We observe the ranking provided by Bing for each search impression (i.e., one user submitting one query to Bing and receiving a set of results), which represents the exact outcome of Bing’s ranking algorithm for that impression.

**Table 3.** User (Normalized) Query-Specific Preferences

Query	1. Watch online	2. Prediction	...	4. Stats	5. Rivalry	...
“live super bowl stream”						
$\theta_q$	0.6701	0.1487	...	0.0230	0.0114	...
$\theta_q + \varepsilon_q$	0.5902	0.0154	...	0.1484	0.0177	...
“facts about the super bowl”						
$\theta_q$	0.0020	0.9162	...	0.0015	0.0733	...
$\theta_q + \varepsilon_q$	0.0397	0.6397	...	0.0408	0.0443	...
“NFL MVPs”						
$\theta_q$	0.0016	0.0075	...	0.9855	0.0006	...
$\theta_q + \varepsilon_q$	0.0084	0.0337	...	0.7897	0.0212	...

*Notes.* The parameter  $\theta_q$  indicates the topic intensities of query  $q$ , and  $\theta_q + \varepsilon_q$  indicates the user click-through preferences underlying query  $q$ . For ease of interpretation, both  $\theta_q$  and  $\theta_q + \varepsilon_q$  are normalized so that the sum across topics is 1.

**Table 4.** User Click-Through Preferences for the Query “Super Bowl” Across Time

Topic	Before	During	After
1. Watch online	1.6623	6.8292	0.0247
2. Prediction	1.5412	0.6810	0.0202
3. Falcons	0.2522	1.0020	1.6623
4. Stats	6.8292	0.0247	0.2325
5. Rivalry	0.6810	0.0202	0.2522
6. Commercial	1.0020	1.6624	6.8292
7. Patriots	0.0247	0.2325	0.6810
8. Halftime show	2.1343	1.0788	1.6448

**5.4.1. Evaluation Metric.** From a practical perspective, users prefer to have the most relevant results at the top of the SERP (Liu et al. 2009). This suggests that search engines should ideally position the clicked link at the top of the SERP. Consider the ranking of results presented by Bing in search impression  $i$  for which a user clicked on one of the results. We assume that Bing presented the (organic) results in decreasing order of predicted CTR. Let  $R_i^{Bing}$  be the position of the link on which the user clicked. For each search impression  $i$ , we also derive the ranking recommended by our model based on CTR predictions. We denote as  $R_i^{model}$  the position that was recommended by our model for the link on which the user clicked. In cases for which  $R_i^{model} < R_i^{Bing}$ , by following our model’s recommendations, the search engine would have been able to rank the link that was preferred by the user even higher, which would have been desirable.

Therefore, comparing  $R_i^{model}$  to  $R_i^{Bing}$  allows us to explore whether our proposed model may potentially help Bing improve its ranking algorithm. This analysis can only be performed on search impressions that resulted in a click, which leaves us with more than one million impressions.<sup>10</sup> We note that such comparison is a conservative test of the ranking generated by the proposed model. Indeed, to the extent that user click behavior is influenced by the actual position of links on the SERP, clicks should be biased toward conforming to the actual positions selected by the search engine (Hotchkiss et al. 2005, Joachims et al. 2005, Varian 2007). For example, in our data, we find that the percentage of search impressions for which  $R_i^{Bing} = 1$  is 54% for annual events and 67% for weekly shows. In such cases, the best our model can do is tie with Bing’s performance.

**Table 5.** Predicting Clicks at the Search Impression Level

Prediction	Weekly shows		Annual events	
	In-sample	Out-sample	In-sample	Out-sample
CTR <sup>1</sup>	0.125	0.121	0.153	0.161
CTR <sup>0</sup>	0.104	0.098	0.107	0.111

**5.4.2. Performance Comparisons.** Table 6 compares, at the search impression level, the actual position of the clicked link ( $R_i^{Bing}$ ) with the suggested position derived from our content-based model ( $R_i^{model}$ ). We again separately report performance for in-sample and out-sample data, for weekly shows and for annual events. The left panel is the average across all the searches for each type of show. We find that for both weekly shows and annual events, our content-based search model performs significantly better than the benchmark for both the in-sample and out-sample datasets. Therefore, the results suggest that Bing could use the proposed model to improve its ranking.

**5.4.3. Moderators of Performance.** We now study potential moderators of the performance of the proposed model. In other words, we identify conditions under which the search engine would be able to detect that the proposed model is likely to be more or less useful.

First, we split the (query, time) pairs on the basis of whether the link with the highest CTR among in-sample search impressions also had the highest average position among all links that were returned by the search engine for that (query, time) combination. For each (query, time) pair, we compute the average position of each link across in-sample search impressions, identify the link with the highest in-sample CTR, and denote as  $\tilde{R}_{in-sample}^{Bing}$  the rank of the average position of the link with the highest in-sample CTR. The middle panel in Table 6 focuses on the 37% of observations for which in-sample  $\tilde{R}_{in-sample}^{Bing} > 1$ . In these situations, the ordering of links proposed by the search engine is not aligned with the observed ranking (implied by aggregated CTR) in the in-sample data, which is more likely to call for an alternative ranking of the links. The rightmost panel focuses on the remaining 63% of observations for which in-sample  $\tilde{R}_{in-sample}^{Bing} = 1$ . In these cases, the in-sample data provides no evidence that the average ordering presented by the search engine is suboptimal for the (query, time) combination. As expected, Bing’s ranking improves when  $\tilde{R}_{in-sample}^{Bing} = 1$ , but our model still significantly outperforms Bing in these cases. That is, even in situations where Bing is expected to perform well on average based on in-sample data, our content-based model can help improve its ranking algorithm.

Next, we investigate some query-level characteristics that may influence our model’s ranking performance. Shorter queries may contain less information about users’ actual search intent, making the ranking problem more difficult. Similarly, if a query is ambiguous, there will be greater uncertainty regarding the type of information for which users are searching, which could also make the ranking problem more challenging. We measure query length as the number of words in the selected vocabulary of



**Table 6.** Improving the Search Engine's Ranking

Data	Ranking	All observations		$\tilde{R}_{\text{in-sample}}^{\text{Bing}} > 1$		$\tilde{R}_{\text{in-sample}}^{\text{Bing}} = 1$	
		Weekly	Annual	Weekly	Annual	Weekly	Annual
In-sample	$R^{\text{model}}$	1.214	1.613	1.282	1.941	1.145	1.524
	$R^{\text{Bing}}$	1.685	1.933	1.806	2.351	1.561	1.820
Out-sample	$R^{\text{model}}$	1.402	1.854	1.461	1.808	1.359	1.870
	$R^{\text{Bing}}$	1.609	1.989	1.728	2.198	1.520	1.917

Note. As the table shows, we find that  $R^{\text{model}}$  is always significantly smaller than  $R^{\text{Bing}}$  ( $p < 0.001$ ), and the difference is larger when  $\tilde{R}_{\text{in-sample}}^{\text{Bing}} > 1$ .

the corpus that appear in the query. We measure query ambiguity as its topic entropy, which captures the uncertainty of a document's topic distribution (Abhishek et al. 2018). More formally, for a query  $q$  with a (normalized) topic distribution  $\theta_q$ , its topic entropy is defined as  $\text{Entropy}_q = -\sum_{k=1}^K \theta_{qk} \log(\theta_{qk})$ . Higher topic entropy means greater ambiguity. We find that for the out-sample data, the average query length is 3.171, with a standard deviation of 1.081, and the average query entropy is  $-4.4513$ , with a standard deviation of 56.8107.

We focus this analysis on the set of search impressions in the out-sample data with at least one click. For each search impression  $i$  of query  $q$  at time  $t$ , the predicted position of any of its search results  $j$  is an integer from 1 to 8 (from the top to the bottom). Hence, we consider an ordered probit regression model in which the dependent variable is  $R_i^{\text{model}}$ , and the key explanatory variables include query length, ambiguity, and their interaction term. We standardize these two variables to make their coefficients comparable. We use fixed effects to control for actual position, TV show, and time. The estimation results are reported in Table 7. All the intercepts and our key independent variables are statistically significant ( $p < 0.01$ ). In particular, the positive coefficient of query length means that holding all other variables constant, when increasing the query length by one unit, the ordered odds of the clicked link having a worse position according to our model is  $\exp(0.0597) = 1.0615$  times greater. The negative coefficient of query ambiguity means that holding all other variables constant, when increasing the query ambiguity by one unit, the ordered odds of the clicked link having a worse position according to our model is  $\exp(0.1123) = 1.1188$  times smaller. Therefore, these results suggest that our model tends to provide better ranking predictions (lower value of  $R_i^{\text{model}}$ ) for shorter and more ambiguous search queries.

## 6. Discussion, Conclusion, and Future Research

In this paper, we develop a model that links the content preferences of search engine users to search

volume and CTR. Our model estimates a latent vector of content preferences over topics revealed by each query, where preferences are allowed to vary systematically along specific dimensions (in our case, before, during, and after a TV show is aired). It also associates the content of each search query and search result description with a latent vector over the same set of topics, constraining both sets of vectors to be sparse and nonnegative. To facilitate efficient and scalable inference, we develop a full Bayesian inference procedure via a variational inference algorithm.

Our proposed model has a novel structure within the broad literature of topic modeling and search, and our paper makes a methodological contribution to this literature. Managerially, we find that our model can generate meaningful substantive output that may lead to important insights for search engines and advertisers. More specifically, our framework can interpret and explain search volume, clicks on links, and the relationship between the two. It can also automatically identify, interpret, and quantify whether and how user content preferences vary across contexts. More importantly, our model can be used to predict CTR even for links for which no prior data are available, and it can help search engines improve their ranking by better accommodating users' context-dependent preferences.

**Table 7.** Ordered Probit Regression of  $R_i^{\text{model}}$ 

Variable	Estimates
Intercept: 1   2	0.8439 (0.0070)
Intercept: 2   3	1.2103 (0.0072)
Intercept: 3   4	1.4182 (0.0074)
Intercept: 4   5	1.5739 (0.0076)
Intercept: 5   6	1.8005 (0.0079)
Intercept: 6   7	2.0906 (0.0085)
Intercept: 7   8	2.4851 (0.0100)
Query length	0.0597 (0.0035)
Query ambiguity	-0.1123 (0.0031)
Query length $\times$ ambiguity	-0.0412 (0.0050)
Bing position fixed effect	Yes
TV show fixed effect	Yes
Time fixed effect	Yes
Number of observations	192,532
Akaike Information Criterion	371,326

Note. Standard errors are reported in parentheses.



Although we focus on organic search results in this paper, as previously mentioned, our model can be easily generalized to sponsored results. In such contexts, our model may not be relevant for optimizing the ranking of sponsored search results, but it may be useful for determining how much an advertiser should be willing to pay to reach a user who submits a particular query in a particular context, and for establishing which ad copy should be shown to this user. For example, our model allows advertisers to quantify how well the copy of a given ad is aligned with the content preferences of users who submitted a particular query in a particular context, even if the ad copy has not been previously tested. This information can help advertisers decide which ad copy should be shown to which users in which context and becomes particularly valuable when advertisers rely on moment marketing companies (e.g., TVTY) to instantly launch and optimize search advertising campaigns triggered by offline events such as TV commercials (Liu and Hill 2021). In addition, search engines now provide advertisers detailed campaign performance data across many contexts (such as time, geographical location, and some demographics). Variants of our model can help advertisers extract meaningful insights from such unstructured data and guide their decision making.

We close by highlighting additional areas for future research. First, future research could include field experiments that precisely measure the impact of implementing the proposed model on the search engine's key performance indicators. Second, our proposed modeling framework can be easily generalized to study other dimensions that may describe the search context, such as location, device, and demographics. Third, whereas in this paper we model aggregate users' search behavior, future research could extend our model with an individual-specific (heterogeneous) preference structure, especially for panel data in contexts like social networking and online shopping. Finally, our model could be applied to any search engine that involves querying and clicking behavior, such as those embedded within YouTube, Amazon, and Priceline. As such, our model could be generalized to online communities to understand users' readership and interests as well as the structure of online content. Ultimately, it can be used by sites to improve their content and recommend content/friends to their users while also addressing the cold-start problem and capturing systematic variations in preferences across contexts.

## Acknowledgments

This research was done when S. Hill was a senior principal researcher at Microsoft Research. This paper is based on the third chapter of J. Liu's dissertation work. The authors

thank Maxime Cohen and Hema Yoganasimhan for their comments on this paper.

## Appendix A. Complete Conditionals

To facilitate inference, we first augment the model with auxiliary variables, which makes it conditionally conjugate. These auxiliary variables are added for Equations (1)–(4). More specifically, for each query  $q$  in Equation (1), we add  $K$  latent variables for each word  $v$ ,  $d_{qv}$ , which denotes the number of words  $v$  in query  $q$  that belong to topic  $k$ . Thus,  $d_{qv} \sim \text{Poisson}(\theta_{qk}\phi_{vk})$ , which are integers such that the word count  $w_{qv} = \sum_k d_{qv}$ . Similarly, for each link  $p$  in Equation (2), we add  $K$  latent variables for each word  $v$ :  $x_{pv} \sim \text{Poisson}(\theta_{pk}\phi_{vk})$ , which are integers such that the word count  $w_{pv} = \sum_k x_{pv}$ . For query  $q$  at time  $t$  in Equation (3), we add  $K$  latent variables:  $y_{qt}^a \sim \text{Poisson}(\theta_{qk}\beta_{tk})$ ,  $K$  latent variables:  $y_{qt}^b \sim \text{Poisson}(\theta_{qk}\gamma_{tk})$ , and one latent variable:  $y_{qt}^c \sim \text{Poisson}(\alpha_q)$ , so that their summation satisfies  $S_{qt} = \sum_k (y_{qt}^a + y_{qt}^b) + y_{qt}^c$ . For each click-through count  $C_{pqt}$  in Equation (4), we add  $K$  latent variables:  $z_{pqt}^a \sim \text{Poisson}(\theta_{pk}\theta_{qk})$ ,  $K$  latent variables:  $z_{pqt}^b \sim \text{Poisson}(\theta_{pk}\epsilon_{qk})$ ,  $K$  latent variables:  $z_{pqt}^c \sim \text{Poisson}(\theta_{pk}\beta_{tk})$ , and one latent variable:  $z_{pqt}^d \sim \text{Poisson}(\lambda_{X_{pqt}})$ , such that  $C_{pqt} = \sum_k (z_{pqt}^a + z_{pqt}^b + z_{pqt}^c) + z_{pqt}^d$ . A sum of independent Poisson random variables is itself a Poisson with a rate equal to the sum of the rates, so these new latent variables preserve the marginal distribution of the observations:

$$\begin{aligned} L(\Theta) = \sum_q & \left\{ \log \Pr(\alpha_q) + \log \Pr(\theta_q) + \log \Pr(\epsilon_q) \right. \\ & + \sum_v [\log \Pr(w_{qv}|d_{qv}) + \log \Pr(d_{qv}|\theta_q, \phi_v)] \\ & + \sum_t [\log \Pr(S_{qt}|y_{qt}) + \log \Pr(y_{qt}|\theta_q, \alpha_q, \beta_t, \gamma_t)] \\ & + \sum_{p,t} [\log \Pr(C_{pqt}|z_{pqt}) + \log \Pr(z_{pqt}|\lambda, \epsilon_q, \beta_t, \theta_q, \theta_p)] \Big\} \\ & + \sum_p \left\{ \log \Pr(\theta_p) + \sum_v [\log \Pr(w_{pv}|x_{pv}) \right. \\ & + \log \Pr(x_{pv}|\theta_p, \phi_v)] \Big\} \\ & + \sum_t (\log \Pr(\beta_t) + \log \Pr(\gamma_t)) \\ & + \sum_v \log \Pr(\phi_v) + \sum_j \log \Pr(\lambda_j). \end{aligned} \quad (\text{A.1})$$

After the data augmentation, the parameters that we need to estimate include:

$$\Theta = \{\phi_k, \theta_q, \theta_p, \epsilon_q, \gamma_t, \beta_t, \alpha_q, \lambda, d_{qv}, x_{pv}, y_{qt}, z_{pqt}\}.$$

Note that when the word count, query usage, or click-through is zero, these auxiliary variables are not random—the posterior distribution will place all its mass on the zero vector. Therefore, the inference procedure will only iterate over nonzero observations (i.e.,  $\{w_{qv} > 0, w_{pv} > 0, S_{qt} > 0, C_{pqt} > 0\}$ ). The full joint log-likelihood function of the data and these parameters can be expressed as Equation (A.1). In the

following, we present the complete conditionals of all these parameters.

### A.1. Topics

The prior of the topic  $\phi_{vk}$  has gamma prior, and its data  $\{d_{qv}k\}_q$  and  $\{x_{pv}k\}_p$  all have Poisson distribution. Hence, the complete conditional of  $\phi_{vk}$  is conjugate and also gamma:

$$\phi_{vk} | \{d_{qv}k, \theta_{qk}\}_q, \{x_{pv}k, \theta_{pk}\}_p \sim \text{Gamma} \left( a_1 + \sum_q d_{qv}k + \sum_p x_{pv}k, a_2 + \sum_q \theta_{qk} + \sum_p \theta_{pk} \right). \quad (\text{A.2})$$

### A.2. Topic Intensities

For each query  $q$ , the prior distribution of  $\theta_{qk}$  is gamma, and its data  $\{d_{qv}k\}_v$ ,  $\{y_{qtk}^a, y_{qtk}^b\}_t$ , and  $\{z_{pqtk}^a\}_{p,t}$  all follow Poisson. Therefore, the posterior distribution of  $\theta_{qk}$  is also gamma:

$$\begin{aligned} \theta_{qk} | \{d_{qv}k, \phi_{vk}\}_v, \{y_{qtk}^a, \beta_t, y_{qtk}^b, \gamma_t\}_t, \{z_{pqtk}^a, \theta_p\}_{p,t} \\ \sim \text{Gamma} \left( c_1 + \sum_v d_{qv}k + \sum_t (y_{qtk}^a + y_{qtk}^b) \right. \\ \left. + \sum_{p,t} z_{pqtk}^a, c_2 + \sum_v \phi_{vk} + \sum_t (\beta_t + \gamma_t) + \sum_{p,t} \theta_{pk} \right). \quad (\text{A.3}) \end{aligned}$$

It is clear that the model learns the topic distribution of a search query by combining the textual information with user query usage as well as click-through data. Thus, the model can naturally address the sparsity of the textual information in user search queries (Liu and Toubia 2018), hence leading to better interpretability through a semi-supervised learning approach. The supervision is guided by how users react to this textual information. Following the same argument, for each link  $p$ , its intensity for topic  $k$  satisfies:

$$\begin{aligned} \theta_{pk} | \{x_{pv}k, \phi_{vk}\}_v, \{z_{pqtk}^a, z_{pqtk}^b, z_{pqtk}^c, \theta_{qk}, \varepsilon_{qk}, \beta_{tk}\}_{q,t} \\ \sim \text{Gamma} \left( b_1 + \sum_v x_{pv}k + \sum_{q,t} (z_{pqtk}^a + z_{pqtk}^b + z_{pqtk}^c), \right. \\ \left. b_2 + \sum_v \phi_{vk} + \sum_{q,t} (\theta_{qk} + \varepsilon_{qk} + \beta_{tk}) \right). \quad (\text{A.4}) \end{aligned}$$

In this case, the model learns the topic distribution of a link by combining its textual information with user click-through data, which can also be understood as semi-supervised learning of link description.

### A.3. User Preferences

We first derive the full conditional of user time-specific preference for a topic  $\beta_{tk}$ , whose data include  $\{y_{qtk}^a\}_q$  and  $\{z_{pqtk}^c\}_{p,q}$ :

$$\begin{aligned} \beta_{tk} | \{\theta_q, y_{qtk}^a\}_q, \{\theta_{pk}, z_{pqtk}^c\}_{p,q} \sim \text{Gamma} \\ \left( e_1 + \sum_q y_{qtk}^a + \sum_{p,q} z_{pqtk}^c, e_2 + \sum_q \theta_{qk} + \sum_{p,q} \theta_{pk} \right). \quad (\text{A.5}) \end{aligned}$$

It is apparent that in posterior, the model learns user time-specific content preferences through both query usage and subsequent click throughs. This inference approach is consistent with Liu and Toubia (2018, 2020), who suggested that queries may not be a direct representation of user preferences and that it is important to link queries to their search results to better estimate user preferences.

Similarly, we derive the full conditional of the offset of user time-specific preferences  $\gamma_t$ , whose data include only  $\{y_{qtk}^b\}_q$ , as:

$$\gamma_t | \{\theta_q, y_{qtk}^b\}_q \sim \text{Gamma} \left( e_1 + \sum_q y_{qtk}^b, e_2 + \sum_q \theta_{qk} \right). \quad (\text{A.6})$$

That is, the model learns the topic offset of user time-specific preferences through the topic intensities of the queries that users are mostly likely to issue. Finally, the data for the offset of user query-specific preferences  $\varepsilon_q$  include  $\{z_{pqtk}^b\}_p$ , which all follow a Poisson distribution. Thereby, the complete conditional of  $\varepsilon_q$  is:

$$\varepsilon_q | \{\theta_{pk}, z_{pqtk}^b\}_{p,t} \sim \text{Gamma} \left( e_1 + \sum_{p,t} z_{pqtk}^b, e_2 + \sum_{p,t} \theta_{pk} \right). \quad (\text{A.7})$$

That is, the model learns the topic offset of user query-specific preferences through the topic intensities of the links that users tend to click.

### A.4. Auxiliary Poisson Variables

The auxiliary variable  $d_{qv}$  for the word count in a query is a  $K$ -dimensional latent vector of Poisson count. When conditional on the observed sum,  $w_{qv}$ ,  $d_{qv}$  is distributed as a multinomial for which the parameter is the normalized set of rates (Cemgil 2009). Therefore, the complete conditional of  $d_{qv}$  is:

$$d_{qv} | w_{qv}, \theta_q, \phi_v \sim \text{Mult} \left( w_{qv}, \frac{\theta_q \cdot \phi_v}{\theta_q^T \phi_v} \right), \quad (\text{A.8})$$

where  $\cdot$  denotes the element-wise product operation. Likewise, the complete conditional of  $x_{pv}$  for the word count of link  $p$  is a  $K$ -dimensional multinomial whose observed summation is  $w_{pv}$ :

$$x_{pv} | w_{pv}, \theta_p, \phi_v \sim \text{Mult} \left( w_{pv}, \frac{\theta_p \cdot \phi_v}{\theta_p^T \phi_v} \right). \quad (\text{A.9})$$

The complete conditional of  $y_{qt} = (y_{qt}^a, y_{qt}^b, y_{qt}^c)$  for the usage of query  $q$  at time  $t$  is a  $(2K+1)$ -dimensional multinomial whose observed summation is  $S_{qt}$ :

$$y_{qt} | \alpha_q, \theta_q, \beta_t, \gamma_t \sim \text{Mult} \left( S_{qt}, \frac{\theta_q \cdot \beta_t}{N_{qt}}, \frac{\theta_q \cdot \gamma_t}{N_{qt}}, \alpha_q \right), \quad (\text{A.10})$$

where  $N_{qt} = \theta_q^T (\beta_t + \gamma_t) + \alpha_q$ . Lastly, the complete conditional for the click-through count  $z_{pqtk} = (z_{pqtk}^a, z_{pqtk}^b, z_{pqtk}^c, z_{pqtk}^d)$

is a  $(3K + 1)$ -dimensional multinomial whose summation is  $C_{pqt}$ :

$$z_{pqt} | \theta_p, \theta_q, \varepsilon_q, \beta_t, \lambda \sim \text{Mult} \left( C_{pqt}, \frac{\theta_p \cdot \theta_q}{N_{pqt}}, \frac{\theta_p \cdot \varepsilon_q}{N_{pqt}}, \frac{\theta_p \cdot \beta_t}{N_{pqt}}, \frac{\lambda_{X_{pqt}^{\text{pos}}}}{N_{pqt}} \right), \quad (\text{A.11})$$

where  $N_{pqt}$  denotes the summation  $\theta_p^T (\theta_q + \varepsilon_q + \beta_t) + \lambda_{X_{pqt}^{\text{pos}}}$ .

### A.5. Fixed Position Effect

Suppose the (rounded) observed position on the SERP is  $X_{pqt}^{\text{pos}} \in \{1, 2, \dots, N\}$ . We assume that the prior distribution of each  $N$  position effect  $\lambda_j$  is Gamma( $h_1, h_2$ ). Given that the data for these scale parameters are  $\{z_{pqt}^d, X_{pqt}^{\text{pos}}\}_{p,q,t}$ , which follows a Poisson distribution, the complete conditional for  $\lambda_j$  follows a gamma distribution:

$$\lambda_j | \{z_{pqt}^d, X_{pqt}^{\text{pos}}\}_{p,q,t} \sim \text{Gamma} \left( h_1 + \sum_{q,p,t} z_{pqt}^d I(X_{pqt}^{\text{pos}} = j), h_2 + \sum_{q,p,t} I(X_{pqt}^{\text{pos}} = j) \right). \quad (\text{A.12})$$

### A.6. Search Volume Baseline

The posterior distribution of the baseline search volume of query  $\alpha_q$  is conjugate with gamma prior and Poisson data  $\{y_{qt}^c\}$ , such that:

$$\alpha_q | \{y_{qt}^c\}_t \sim \text{Gamma} \left( e_1 + \sum_t y_{qt}^c, e_2 + T \right). \quad (\text{A.13})$$

## Appendix B. Variational Inference Algorithm

### B.1. Variational Family

The first step is to define a variational distribution  $f(\cdot)$  of each model variable as an approximation of the actual

posterior distribution that is given in Appendix A. In our application, we define the variational family  $f(\varphi, \theta, \beta, \varepsilon, \alpha, \lambda, d, x, y, z)$  with a mean-field variational form (Jordan et al. 1999) in which all of the variables are assumed to be independent, and each is governed by its own distribution:

$$\begin{aligned} f(\varphi, \theta, \beta, \varepsilon, \alpha, \lambda, d, x, y, z) &= \left[ \sum_j f(\lambda_j) \right] \left[ \prod_{v,k} f(\varphi_{vk}) \right] \left[ \prod_{q,k} f(\theta_{qk}) f(\varepsilon_{qk}) \right] \\ &\times \left[ \prod_{p,k} f(\theta_{pk}) \right] \left[ \prod_q f(\alpha_q) \right] \\ &\times \left[ \prod_{t,k} f(\beta_{tk}) f(\gamma_{tk}) \right] \left[ \prod_{q,v,k} f(d_{qv}) \right] \left[ \prod_{p,v,k} f(x_{pvk}) \right] \\ &\times \left[ \prod_{q,t,k} f(y_{qt}) \right] \left[ \prod_{p,q,t,k} f(z_{pqt}) \right]. \end{aligned} \quad (\text{B.1})$$

Table B.1 summarizes the complete conditionals derived in Appendix A and the corresponding variational parameters for all the variables for which the parameter form of multinomial distribution corresponds to its formula in an exponential family. For example, the variational distribution for preference  $\varepsilon_{qk}$  is Gamma( $\tilde{\varepsilon}_{qk}^{\text{shp}}, \tilde{\varepsilon}_{qk}^{\text{rte}}$ ). We denote the shape with the superscript “shp” and rate with the superscript “rte.” The variational factors for these auxiliary variables are all multinomial, which are the same as their conditional distribution in Appendix A.

### B.2. Variational Objective Function

Variational inference minimizes the Kullback-Leibler divergence from the variational distribution to the posterior distribution. This is equivalent to maximizing the evidence lower bound (ELBO), a lower bound on the log-likelihood

**Table B.1.** Complete Conditionals and Variational Parameters

Variable	Type	Complete conditional	Variational parameter
$\varphi_{vk}$	Gamma	$a_1 + \sum_q d_{qv} + \sum_p x_{pvk}, a_2 + \sum_q \theta_{qk} + \sum_p \theta_{pk}$	$\tilde{\varphi}_{vk}^{\text{shp}}, \tilde{\varphi}_{vk}^{\text{rte}}$
$\theta_{qk}$	Gamma	$c_1 + \sum_v d_{qv} + \sum_t (y_{qt}^a + y_{qt}^b) + \sum_{p,t} z_{pqt}^a,$ $c_2 + \sum_v \varphi_{vk} + \sum_t (\beta_{tk} + \gamma_{tk}) + \sum_{p,t} \theta_{pk}$	$\tilde{\theta}_{qk}^{\text{shp}}, \tilde{\theta}_{qk}^{\text{rte}}$
$\theta_{pk}$	Gamma	$b_1 + \sum_v x_{pvk} + \sum_{q,t} (z_{pqt}^a + z_{pqt}^b + z_{pqt}^c),$ $b_2 + \sum_v \varphi_{vk} + \sum_{q,t} (\theta_{qk} + \varepsilon_{qk} + \beta_{tk})$	$\tilde{\theta}_{pk}^{\text{shp}}, \tilde{\theta}_{pk}^{\text{rte}}$
$\varepsilon_{qk}$	Gamma	$e_1 + \sum_{p,t} z_{pqt}^b, e_2 + \sum_{p,t} \theta_{pk}$	$\tilde{\varepsilon}_{qk}^{\text{shp}}, \tilde{\varepsilon}_{qk}^{\text{rte}}$
$\beta_{tk}$	Gamma	$e_1 + \sum_q y_{qt}^a + \sum_p z_{pqt}^c, e_2 + \sum_q \theta_{qk} + \sum_p \theta_{pk}$	$\tilde{\beta}_{tk}^{\text{shp}}, \tilde{\beta}_{tk}^{\text{rte}}$
$\gamma_{tk}$	Gamma	$e_1 + \sum_q y_{qt}^b, e_2 + \sum_q \theta_{qk}$	$\tilde{\gamma}_{tk}^{\text{shp}}, \tilde{\gamma}_{tk}^{\text{rte}}$
$\alpha_q$	Gamma	$e_1 + \sum_t y_{qt}^c, e_2 + T$	$\tilde{\alpha}_q^{\text{shp}}$
$\lambda_j$	Gamma	$h_1 + \sum_{q,p,t} z_{pqt}^d I(X_{pqt}^{\text{pos}} = j), h_2 + \sum_{q,p,t} I(X_{pqt}^{\text{pos}} = j)$	$\tilde{\lambda}_j^{\text{shp}}$
$d_{qv}$	Multinomial	$\log \theta_{qk} + \log \varphi_{vk}$	$\mu_{qv}^d$
$x_{pvk}$	Multinomial	$\log \theta_{pk} + \log \varphi_{vk}$	$\mu_{pvk}^x$
$y_{qt}$	Multinomial	$\log \theta_{qk} + \log \beta_{tk}, \text{ if } 1 \leq k \leq K;$ $\log \theta_{qk} + \log \gamma_{tk}, \text{ if } K+1 \leq k \leq 2K;$ $\log \alpha_q, \text{ if } k = 2K+1$	$\mu_{qt}^y$
$z_{pqt}$	Multinomial	$\log \theta_{pk} + \log \theta_{qk}, \text{ if } 1 \leq k \leq K;$ $\log \theta_{pk} + \log \varepsilon_{qk}, \text{ if } K+1 \leq k \leq 2K;$ $\log \theta_{pk} + \log \beta_{tk}, \text{ if } 2K+1 \leq k \leq 3K;$ $\log \lambda_{X_{pqt}^{\text{pos}}}, \text{ if } k = 3K+1$	$\mu_{pqt}^z$

of the data (Jordan et al. 1999, Hoffman et al. 2013). Based on the full joint log-likelihood function in Equation (A.1) and the variational family in Equation (B.1), the ELBO of the full model can be written as:

$$L(f) = E_f[\log \Pr(\boldsymbol{\varphi}, \boldsymbol{\theta}, \beta, \varepsilon, \gamma, \alpha, \lambda, d, x, y, z)] - E_f[\log f(\boldsymbol{\varphi}, \boldsymbol{\theta}, \beta, \varepsilon, \gamma, \alpha, \lambda, d, x, y, z)]. \quad (\text{B.2})$$

In the first line, the first term is the expected log joint distribution of the model, and the second term is the entropy of the variational distribution, which encourages distribution to be spread across configurations.

A few expectation terms in Equation (B.2) do not have closed-form solutions and hence are intractable. Fortunately, these difficult terms appear in both the joint log-likelihood function and the entropy and thus can be canceled. To compute the expectation of a logarithm variable  $x$  with a prior  $f(\cdot) \sim \text{Gamma}(a, b)$ , we use the fact that  $E_f[\log x] = \Psi(a) - \log b$ , where  $\Psi(\cdot)$  denotes the digamma function (the first derivative of the log gamma function). After replacing all the expectation terms of these latent variables under variational distribution  $f(\cdot)$  and then removing terms that are constant, we obtain a surrogate expression of the ELBO in Equation (B.3) that depends only on data, variational parameters, and hyperparameters:

$$\begin{aligned} L(f) \propto & \sum_j \left\{ \left( h_1 - \tilde{\lambda}_j^{shp} \right) \Psi \left( \tilde{\lambda}_j^{shp} \right) + \frac{\lambda_j^{rte} - h_2}{\lambda_j^{rte}} \tilde{\lambda}_j^{shp} + \log \gamma \left( \tilde{\lambda}_j^{shp} \right) \right\} \\ & + \sum_q \left\{ \left( e_1 - \tilde{\alpha}_q^{shp} \right) \Psi \left( \tilde{\alpha}_q^{shp} \right) + \frac{\alpha_q^{rte} - h_2}{\alpha_q^{rte}} \tilde{\alpha}_q^{shp} + \log \gamma \left( \tilde{\alpha}_q^{shp} \right) \right\} \\ & + \sum_{v,k} \left\{ \left( a_1 - \tilde{\boldsymbol{\varphi}}_{vk}^{shp} \right) \Psi \left( \tilde{\boldsymbol{\varphi}}_{vk}^{shp} \right) - a_1 \log \tilde{\boldsymbol{\varphi}}_{vk}^{rte} \right. \\ & \left. - a_2 \tilde{\boldsymbol{\varphi}}_{vk}^{shp} / \tilde{\boldsymbol{\varphi}}_{vk}^{rte} + \tilde{\boldsymbol{\varphi}}_{vk}^{shp} + \log \gamma \left( \tilde{\boldsymbol{\varphi}}_{vk}^{shp} \right) \right\} \\ & + \sum_{p,k} \left\{ \left( b_1 - \tilde{\boldsymbol{\theta}}_{pk}^{shp} \right) \Psi \left( \tilde{\boldsymbol{\theta}}_{pk}^{shp} \right) - b_1 \log \tilde{\boldsymbol{\theta}}_{pk}^{rte} + \right. \\ & \left. - b_2 \tilde{\boldsymbol{\theta}}_{pk}^{shp} / \tilde{\boldsymbol{\theta}}_{pk}^{rte} + \tilde{\boldsymbol{\theta}}_{pk}^{shp} + \log \gamma \left( \tilde{\boldsymbol{\theta}}_{pk}^{shp} \right) \right\} \\ & + \sum_{q,k} \left\{ \left( c_1 - \tilde{\boldsymbol{\theta}}_{qk}^{shp} \right) \Psi \left( \tilde{\boldsymbol{\theta}}_{qk}^{shp} \right) - c_1 \log \tilde{\boldsymbol{\theta}}_{qk}^{rte} \right. \\ & \left. - b_2 \tilde{\boldsymbol{\theta}}_{qk}^{shp} / \tilde{\boldsymbol{\theta}}_{qk}^{rte} + \tilde{\boldsymbol{\theta}}_{qk}^{shp} + \log \gamma \left( \tilde{\boldsymbol{\theta}}_{qk}^{shp} \right) \right\} \\ & + \left( e_1 - \tilde{\varepsilon}_{qk}^{shp} \right) \Psi \left( \tilde{\varepsilon}_{qk}^{shp} \right) - e_1 \log \tilde{\varepsilon}_{qk}^{rte} \\ & \left. - e_2 \tilde{\varepsilon}_{qk}^{shp} / \tilde{\varepsilon}_{qk}^{rte} + \tilde{\varepsilon}_{qk}^{shp} + \log \gamma \left( \tilde{\varepsilon}_{qk}^{shp} \right) \right\} \\ & + \sum_{t,k} \left\{ \left( e_1 - \tilde{\beta}_{tk}^{shp} \right) \Psi \left( \tilde{\beta}_{tk}^{shp} \right) - e_1 \log \tilde{\beta}_{tk}^{rte} \right. \\ & \left. - e_2 \tilde{\beta}_{tk}^{shp} / \tilde{\beta}_{tk}^{rte} + \tilde{\beta}_{tk}^{shp} + \log \gamma \left( \tilde{\beta}_{tk}^{shp} \right) \right\} \\ & + \left( e_1 - \tilde{\gamma}_{tk}^{shp} \right) \Psi \left( \tilde{\gamma}_{tk}^{shp} \right) - e_1 \log \tilde{\gamma}_{tk}^{rte} - e_2 \tilde{\gamma}_{tk}^{shp} / \tilde{\gamma}_{tk}^{rte} \\ & \left. + \tilde{\gamma}_{tk}^{shp} + \log \gamma \left( \tilde{\gamma}_{tk}^{shp} \right) \right\} \\ & + \sum_{q,v,k} \left\{ - \frac{\tilde{\boldsymbol{\theta}}_{qk}^{shp} \tilde{\boldsymbol{\varphi}}_{vk}^{shp}}{\tilde{\boldsymbol{\theta}}_{qk}^{rte} \tilde{\boldsymbol{\varphi}}_{vk}^{rte}} + w_{qv} \mu_{qv}^d \left( \Psi \left( \tilde{\boldsymbol{\theta}}_{qk}^{shp} \right) - \log \tilde{\boldsymbol{\theta}}_{qk}^{rte} \right. \right. \\ & \left. \left. + \Psi \left( \tilde{\boldsymbol{\varphi}}_{vk}^{shp} \right) - \log \tilde{\boldsymbol{\varphi}}_{vk}^{rte} - \log \mu_{qv}^d \right) \right\} \\ & + \sum_{p,v,k} \left\{ - \frac{\tilde{\boldsymbol{\theta}}_{pk}^{shp} \tilde{\boldsymbol{\varphi}}_{vk}^{shp}}{\tilde{\boldsymbol{\theta}}_{pk}^{rte} \tilde{\boldsymbol{\varphi}}_{vk}^{rte}} + w_{pv} \mu_{pv}^x \left( \Psi \left( \tilde{\boldsymbol{\theta}}_{pk}^{shp} \right) - \log \tilde{\boldsymbol{\theta}}_{pk}^{rte} \right. \right. \\ & \left. \left. + \Psi \left( \tilde{\boldsymbol{\varphi}}_{vk}^{shp} \right) - \log \tilde{\boldsymbol{\varphi}}_{vk}^{rte} - \log \mu_{pv}^x \right) \right\} \\ & + \sum_{q,t} \left\{ \sum_{k=1}^K \left[ - \frac{\tilde{\boldsymbol{\theta}}_{qk}^{shp} \tilde{\beta}_{tk}^{shp}}{\tilde{\boldsymbol{\theta}}_{qk}^{rte} \tilde{\beta}_{tk}^{rte}} + S_{qt} \mu_{qt}^y \left( \Psi \left( \tilde{\boldsymbol{\theta}}_{qk}^{shp} \right) \right. \right. \right. \\ & \left. \left. - \log \tilde{\boldsymbol{\theta}}_{qk}^{rte} + \Psi \left( \tilde{\beta}_{tk}^{shp} \right) - \log \tilde{\beta}_{tk}^{rte} - \log \mu_{qt}^y \right) \right] \\ & + \sum_{k=K+1}^{2K} \left[ - \frac{\tilde{\boldsymbol{\theta}}_{qk}^{shp} \tilde{\gamma}_{tk}^{shp}}{\tilde{\boldsymbol{\theta}}_{qk}^{rte} \tilde{\gamma}_{tk}^{rte}} + S_{qt} \mu_{qt}^y \left( \Psi \left( \tilde{\boldsymbol{\theta}}_{qk}^{shp} \right) \right. \right. \\ & \left. \left. - \log \tilde{\boldsymbol{\theta}}_{qk}^{rte} + \Psi \left( \tilde{\gamma}_{tk}^{shp} \right) - \log \tilde{\gamma}_{tk}^{rte} - \log \mu_{qt}^y \right) \right] \\ & \left. - \frac{\tilde{\alpha}_q^{shp}}{\alpha_q^{rte}} + S_{qt} \mu_{qt,2K+1}^y \left( \Psi \left( \tilde{\alpha}_q^{shp} \right) - \log \alpha_q^{rte} - \log \mu_{qt,2K+1}^y \right) \right\} \\ & + \sum_{p,q,t} \left\{ \sum_{k=1}^K \left[ - \frac{\tilde{\boldsymbol{\theta}}_{pk}^{shp} \tilde{\boldsymbol{\theta}}_{qk}^{shp}}{\tilde{\boldsymbol{\theta}}_{pk}^{rte} \tilde{\boldsymbol{\theta}}_{qk}^{rte}} + C_{pqt} \mu_{pqt}^z \left( \Psi \left( \tilde{\boldsymbol{\theta}}_{pk}^{shp} \right) \right. \right. \right. \\ & \left. \left. - \log \tilde{\boldsymbol{\theta}}_{pk}^{rte} + \Psi \left( \tilde{\boldsymbol{\theta}}_{qk}^{shp} \right) - \log \tilde{\boldsymbol{\theta}}_{qk}^{rte} - \log \mu_{pqt}^z \right) \right] \\ & + \sum_{k=K+1}^{2K} \left[ - \frac{\tilde{\boldsymbol{\theta}}_{pk}^{shp} \tilde{\varepsilon}_{qk}^{shp}}{\tilde{\boldsymbol{\theta}}_{pk}^{rte} \tilde{\varepsilon}_{qk}^{rte}} + C_{pqt} \mu_{pqt}^z \left( \Psi \left( \tilde{\boldsymbol{\theta}}_{pk}^{shp} \right) \right. \right. \\ & \left. \left. - \log \tilde{\boldsymbol{\theta}}_{pk}^{rte} + \Psi \left( \tilde{\varepsilon}_{qk}^{shp} \right) - \log \tilde{\varepsilon}_{qk}^{rte} - \log \mu_{pqt}^z \right) \right] \\ & + \sum_{k=2K+1}^{3K} \left[ - \frac{\tilde{\boldsymbol{\theta}}_{pk}^{shp} \tilde{\beta}_{tk}^{shp}}{\tilde{\boldsymbol{\theta}}_{pk}^{rte} \tilde{\beta}_{tk}^{rte}} + C_{pqt} \mu_{pqt}^z \left( \Psi \left( \tilde{\boldsymbol{\theta}}_{pk}^{shp} \right) \right. \right. \\ & \left. \left. - \log \tilde{\boldsymbol{\theta}}_{pk}^{rte} + \Psi \left( \tilde{\beta}_{tk}^{shp} \right) - \log \tilde{\beta}_{tk}^{rte} - \log \mu_{pqt}^z \right) \right] \\ & \left. - \frac{\tilde{\lambda}_{X_{pqt}^{pos}}^{shp}}{\lambda_{X_{pqt}^{pos}}^{rte}} + C_{pqt} \mu_{pqt,3K+1}^z \left( \Psi \left( \tilde{\lambda}_{X_{pqt}^{pos}}^{shp} \right) \right. \right. \\ & \left. \left. - \log \lambda_{X_{pqt}^{pos}}^{rte} - \log \mu_{pqt,3K+1}^z \right) \right\}. \quad (\text{B.3}) \end{aligned}$$

### B.3. Coordinate Ascent Algorithm

#### Algorithm B.1. CAVI for Content-Based Search Model

**Input:** Data  $\{w_q, w_p, S_{qt}, C_{pqt}, X_{pqt}^{pos}\}$ , and tolerance  $\delta$

- 1: Initialize all variational parameters,  $ELBO_0$ ,  $ELBO_1$ , and  $\Delta ELBO$
- 2: **while**  $\Delta ELBO > \delta$  **do**
- 3:  $ELBO_0 = ELBO_1$
- 4: **procedure** Updating Variational Parameters



$$\begin{aligned}
 5: \quad & \tilde{\theta}_{vk}^{shp} = a_1 + \sum_q w_{qv} \mu_{qv}^d + \sum_p w_{pv} \mu_{pv}^x, \tilde{\theta}_{vk}^{rte} = a_2 + \sum_q \frac{\tilde{\theta}_{qk}^{shp}}{\tilde{\theta}_{qk}^{rte}} \\
 & + \sum_p \frac{\tilde{\theta}_{pk}^{shp}}{\tilde{\theta}_{pk}^{rte}} \\
 6: \quad & \tilde{\theta}_{qk}^{shp} = c_1 + \sum_v w_{qv} \mu_{qv}^d + \sum_t S_{qt} (\mu_{qtk}^y + \mu_{qtk}^y) + \sum_{p,t} C_{pqt} \mu_{pqt}^z \\
 7: \quad & \tilde{\theta}_{qk}^{rte} = c_2 + \sum_v \frac{\tilde{\theta}_{vk}^{shp}}{\tilde{\theta}_{vk}^{rte}} + \sum_t \left( \frac{\tilde{\theta}_{tk}^{shp}}{\tilde{\theta}_{tk}^{rte}} + \frac{\tilde{\gamma}_{tk}^{shp}}{\tilde{\gamma}_{tk}^{rte}} \right) + \sum_{p,t} \frac{\tilde{\theta}_{pk}^{shp}}{\tilde{\theta}_{pk}^{rte}} \\
 8: \quad & \tilde{\theta}_{pk}^{shp} = b_1 + \sum_v w_{pv} \mu_{pv}^x + \sum_{q,t} C_{pqt} (\mu_{pqt}^z + \mu_{pqt,K+k}^z + \mu_{pqt,2K+k}^z) \\
 9: \quad & \tilde{\theta}_{pk}^{rte} = b_2 + \sum_v \frac{\tilde{\theta}_{vk}^{shp}}{\tilde{\theta}_{vk}^{rte}} + \sum_{q,t} \left( \frac{\tilde{\theta}_{qk}^{shp}}{\tilde{\theta}_{qk}^{rte}} + \frac{\tilde{\gamma}_{qk}^{shp}}{\tilde{\gamma}_{qk}^{rte}} + \frac{\tilde{\beta}_{pk}^{shp}}{\tilde{\beta}_{pk}^{rte}} \right) \\
 10: \quad & \tilde{\epsilon}_{qk}^{shp} = e_1 + \sum_p C_{pqt} \mu_{pqt}^z, \tilde{\epsilon}_{qk}^{rte} = e_2 + \sum_p \frac{\tilde{\theta}_{pk}^{shp}}{\tilde{\theta}_{pk}^{rte}} \\
 11: \quad & \tilde{\beta}_{tk}^{shp} = e_1 + \sum_q S_{qt} \mu_{qtk}^y + \sum_p C_{pqt} \mu_{pqt}^z, \tilde{\beta}_{tk}^{rte} = e_2, \\
 & + \sum_q \frac{\tilde{\theta}_{qk}^{shp}}{\tilde{\theta}_{qk}^{rte}} + \sum_p \frac{\tilde{\theta}_{pk}^{shp}}{\tilde{\theta}_{pk}^{rte}} \\
 12: \quad & \tilde{\gamma}_{tk}^{shp} = e_1 + \sum_q S_{qt} \mu_{qtk}^y, \tilde{\gamma}_{tk}^{rte} = e_2 + \sum_q \frac{\tilde{\theta}_{qk}^{shp}}{\tilde{\theta}_{qk}^{rte}} \\
 13: \quad & \tilde{\alpha}_q^{shp} = e_1 + \sum_t S_{qt} \mu_{qtk}^y, \tilde{\lambda}_j^{shp} = h_1 + \sum_{q,p,t} C_{pqt} \mu_{pqt}^z, 3K+1 \\
 & \times I(X_{pqt}^{pos} = j) \\
 14: \quad & \mu_{qv}^d \propto \exp \{ \Psi(\tilde{\theta}_{qk}^{shp}) - \log \tilde{\theta}_{qk}^{rte} + \Psi(\tilde{\theta}_{vk}^{shp}) - \log \tilde{\theta}_{vk}^{rte} \} \\
 15: \quad & \mu_{pv}^x \propto \exp \{ \Psi(\tilde{\theta}_{pk}^{shp}) - \log \tilde{\theta}_{pk}^{rte} + \Psi(\tilde{\theta}_{vk}^{shp}) - \log \tilde{\theta}_{vk}^{rte} \} \\
 & \left\{ \begin{array}{l} \exp \{ \Psi(\tilde{\theta}_{qk}^{shp}) - \log \tilde{\theta}_{qk}^{rte} + \Psi(\tilde{\beta}_{tk}^{shp}) - \log \tilde{\beta}_{tk}^{rte} \} \\ \text{if } k \leq K \\ \exp \{ \Psi(\tilde{\theta}_{qk}^{shp}) - \log \tilde{\theta}_{qk}^{rte} + \Psi(\tilde{\gamma}_{tk}^{shp}) - \log \tilde{\gamma}_{tk}^{rte} \} \\ \text{if } K+1 \leq k \leq 2K \\ \exp \{ \Psi(\tilde{\alpha}_q^{shp}) - \log \alpha_q^{rte} \} \\ \text{if } k = 2K+1 \end{array} \right. \\
 16: \quad & \mu_{qtk}^y \propto \left\{ \begin{array}{l} \exp \{ \Psi(\tilde{\theta}_{qk}^{shp}) - \log \tilde{\theta}_{qk}^{rte} + \Psi(\tilde{\theta}_{pk}^{shp}) - \log \tilde{\theta}_{pk}^{rte} \} \\ \text{if } k \leq K \\ \exp \{ \Psi(\tilde{\theta}_{qk}^{shp}) - \log \tilde{\theta}_{qk}^{rte} + \Psi(\tilde{\epsilon}_{qk}^{shp}) - \log \tilde{\epsilon}_{qk}^{rte} \} \\ \text{if } K+1 \leq k \leq 2K \\ \exp \{ \Psi(\tilde{\theta}_{qk}^{shp}) - \log \tilde{\theta}_{qk}^{rte} + \Psi(\tilde{\beta}_{tk}^{shp}) - \log \tilde{\beta}_{tk}^{rte} \} \\ \text{if } 2K+1 \leq k \leq 3K \\ \exp \{ \Psi(\tilde{\lambda}_{X_{pqt}^{pos}}^{shp}) - \log \lambda_{X_{pqt}^{pos}}^{rte} \} \\ \text{if } k = 3K+1 \end{array} \right. \\
 17: \quad & \mu_{pqt}^z \propto \left\{ \begin{array}{l} \exp \{ \Psi(\tilde{\theta}_{qk}^{shp}) - \log \tilde{\theta}_{qk}^{rte} + \Psi(\tilde{\theta}_{pk}^{shp}) - \log \tilde{\theta}_{pk}^{rte} \} \\ \text{if } k \leq K \\ \exp \{ \Psi(\tilde{\theta}_{qk}^{shp}) - \log \tilde{\theta}_{qk}^{rte} + \Psi(\tilde{\epsilon}_{qk}^{shp}) - \log \tilde{\epsilon}_{qk}^{rte} \} \\ \text{if } K+1 \leq k \leq 2K \\ \exp \{ \Psi(\tilde{\theta}_{qk}^{shp}) - \log \tilde{\theta}_{qk}^{rte} + \Psi(\tilde{\beta}_{tk}^{shp}) - \log \tilde{\beta}_{tk}^{rte} \} \\ \text{if } 2K+1 \leq k \leq 3K \\ \exp \{ \Psi(\tilde{\lambda}_{X_{pqt}^{pos}}^{shp}) - \log \lambda_{X_{pqt}^{pos}}^{rte} \} \\ \text{if } k = 3K+1 \end{array} \right. \\
 18: \quad & \text{end procedure} \\
 19: \quad & \text{procedure Updating Metrics for Convergence Evaluation} \\
 20: \quad & \text{Compute } ELBO_1 \text{ using Equation (B.3)} \\
 21: \quad & \Delta ELBO = ELBO_1 - ELBO_0 \\
 22: \quad & \text{end procedure} \\
 23: \quad & \text{end while}
 \end{aligned}$$

24: Compute the posterior estimate as the ratio of the shape and rate variational parameters

**Output:**  $\{\theta_k, \theta_q, \theta_p, \epsilon_q, \gamma_t, \beta_t, \alpha_q, d_{qv}, x_{pv}, y_{qt}, z_{pqt}, \lambda_j\}$

The mean-field variational family enables the CAVI algorithm (Bishop 2006). The central idea of CAVI is to optimize one variational parameter each time while fixing all others. This algorithm guarantees a local optimum of the ELBO (i.e., the bound on the log probability of the model). CAVI is closely related to the Gibbs sampler: the Gibbs sampler maintains a realization of the latent variable and iteratively samples from each variable's complete conditional, whereas CAVI maintains a mean-field variational distribution and iteratively sets each variable's variational factor using the expected log of the complete conditional. If the complete conditional distribution of a latent variable is an exponential family and its corresponding variational distribution has the same form, then its variational parameters have a closed-form solution using the coordinate ascent algorithm (Hoffman et al. 2013). More specifically, the variational parameter equals the expectation of the

conditional parameter in its corresponding complete conditional distribution. Under the conditionally conjugate augmented model, this property is satisfied. Accordingly, for all the variational parameters whose distribution is gamma, the update formulas are simply the expectation of the conditional posterior derived in Appendix A under variational distribution  $f(\cdot)$ . The update formulas for the multinomial variables are also derived based on the expectation of the complete conditionals in Table B.1, and they should be normalized together to ensure they sum to 1. We present the complete procedure in Algorithm B.1.

## Endnotes

<sup>1</sup> For confidentiality, we normalize the total search volume of each query by dividing it by the total search volume of the query “super bowl 50 highlights,” whose normalized search volume hence is 1 in Figure 1.

<sup>2</sup> We find that the positions of the URLs in Figure 2 are quite stable on Bing across impressions. For this particular query, the ranking of the URLs derived from the average position across impressions is the same in all three time windows. More generally, we received confirmation from the company that at the time of writing this paper, Bing did not systematically adjust the ranking of search results around a TV show's airing time. This is also confirmed by the additional data reported in online Appendix D.

<sup>3</sup> Graphical models merge graph theory and probability theory in a powerful formalism for multivariate statistical modeling. Examples include hidden Markov models, latent Dirichlet allocation, and Kalman filters. Graphical models have three appealing features (Jordan 2004): (1) the relationships among different variables are readable from a graph; (2) the models can define a factorization of the joint probabilistic distribution of these variables; and (3) the inference algorithm is connected with the learning algorithm by focusing on the conditional independence of different random variables while maintaining control over the computational cost associated with the models.

<sup>4</sup> Note that in reality, such a description may change across searches. However, based on the large-scale search data we collected from Bing, we find that within a short time window such as two days, there are minimal changes to each link's description. In our empirical study, we use the most frequent version of the description of each link recorded within 24 hours of show time.

<sup>5</sup> One can allow the hyper-prior parameters in the gamma distribution to differ across model parameters. Here, we use the same hyper-prior parameters for different parameters for simplicity of notation. Based on our empirical study, we also find that hyper-prior parameters do not significantly change the estimation results.

<sup>6</sup> One could consider alternative specifications of the position effect as long as the model remains conjugate. For instance, we considered a parametric specification of the position effect  $\lambda/X_{pqt}^{pos}$ , in which case we also allowed the coefficient  $\lambda$  to vary across queries. Although this can capture query heterogeneity, we find that with our corpora, the overall model performance is worse than under a fixed-effect specification, especially for annual TV events.

<sup>7</sup> We set a higher cutoff point for the annual events because they tend to generate much larger search volumes than most weekly TV shows. However, we still have a large number of queries for annual events. See Table 1.

<sup>8</sup> The relevance of word  $w$  to topic  $k$  is measured as  $r(w, k|\pi) = \pi \log(\hat{\phi}_{kw}) + (1 - \pi) \log(\hat{\phi}_{kw}/p_w)$ , where  $\hat{\phi}_{kw}$  is the posterior estimate of the probability of seeing word  $w$  given topic  $k$  (equal to  $\phi_{kw}/\sum_w \phi_{kw}$ ),  $p_w$  is the empirical distribution of word  $w$  in the corpus, and  $\pi$  determines the weight given to the probability of word  $w$  under

topic  $k$  relative to its lift  $\tilde{\phi}_{kw}/p_w$ , both measured on the log scale. We set  $\pi = 0.6$ , following Sievert and Shirley (2014).

<sup>9</sup> The cosine similarity between two vectors  $a$  and  $b$  is computed based on their inner product:  $f(a, b) = a \cdot b / \|a\| \|b\|$ . In our case, it ranges from 0, indicating completely orthogonal, to 1, meaning exactly similar.

<sup>10</sup> Among these selected searches, we find that about 90% of searches generate only one click. Throughout this section, when more than one link on the SERP was clicked, we evaluate only the link with the highest position on Bing (i.e., lower value of  $R$ ). This rule makes our comparison more conservative, as it favors Bing.

## References

- Abhishek V, Gong J, Li B (2018) Examining the impact of contextual ambiguity on search advertising keyword performance: A topic model approach. *MIS Quart.* 42(3):805–829.
- Adomavicius G, Tuzhilin A (2015) *Context-Aware Recommender Systems* (Springer US, Boston), 191–226.
- Agichtein E, Brill E, Dumais S (2006) Improving web search ranking by incorporating user behavior information. *Proc. 29th Annual Internat. ACM SIGIR Conf. Res. Development Inform. Retrieval* (ACM, New York), 19–26.
- Archak N, Ghose A, Ipeirotis PG (2011) Deriving the pricing power of product features by mining consumer reviews. *Management Sci.* 57(8):1485–1509.
- Bischof J, Airoldi EM (2012) Summarizing topical content with word frequency and exclusivity. Langford J, Pineau J, eds. *Proc. 29th Internat. Conf. Machine Learn.* Edinburgh, Scotland, UK (ICML-12), 201–208.
- Bishop CM (2006) *Pattern Recognition and Machine Learning* (Springer-Verlag, New York).
- Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. *J. Machine Learn. Res.* 3:993–1022.
- Büschken J, Allenby GM (2017) Sentence-based text analysis for customer reviews. *Marketing Sci.* 35(6):953–975.
- Canny J (2004) Gap: A factor model for discrete data. *Proc. 27th Annual Internat. ACM SIGIR Conf. Res. Development Inform. Retrieval* (ACM, New York), 122–129.
- Cemgil AT (2009) Bayesian inference for nonnegative matrix factorisation models. *Comput. Intelligence Neurosci.* 2009:785152.
- De los Santos B, Koulayev S (2017) Optimizing click-through in online rankings with endogenous search refinement. *Marketing Sci.* 36(4):471–643.
- Fossen BL, Schweidel DA (2017) Television advertising and online word-of-mouth: An empirical investigation of social TV activity. *Marketing Sci.* 36(1):105–123.
- Ghose A, Ipeirotis PG, Li B (2012) Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Sci.* 31(3):493–520.
- Google (2014) Factors of viewability. Accessed January 2018. <https://www.thinkwithgoogle.com/marketing-resources/data-measurement/5-factors-of-viewability/>.
- Gopalan P, Hofman JM, Blei DM (2013) Scalable recommendation with Poisson factorization. Meila M, Heskes T, eds. *Proc. 31st Conf. Uncertainty Artificial Intelligence* (UAI Press, 2015), 326–335.
- Gopalan PK, Charlin L, Blei D (2014) Content-based recommendations with Poisson factorization. Jordan NI, LeCun Y, Solla SA, eds. *Advances in Neural Information Processing Systems* (MIT Press, Cambridge, MA), 3176–3184.
- Hoffman MD, Blei DM, Wang C, Paisley J (2013) Stochastic variational inference. *J. Machine Learn. Res.* 14(1):1303–1347.
- Hotchkiss G, Alston S, Edwards G (2005) Google eye tracking report: How searchers see and click on Google search results. Report, Enquiro Search Solutions, Kelowna, British Columbia.
- InternetLiveStats (2016) Google search statistics. Accessed January, 2018. <http://www.internetlivestats.com/google-search-statistics/>.
- Joachims T, Granka L, Pan B, Hembrooke H, Gay G (2005) Accurately interpreting clickthrough data as implicit feedback. *Proc. 28th Annual Internat. ACM SIGIR Conf. Res. Development Inform. Retrieval* (ACM, New York), 154–161.
- Jordan MI (2004) Graphical models. *Statist. Sci.* 19(1):140–155.
- Jordan MI, Ghahramani Z, Jaakkola TS, Saul LK (1999) An introduction to variational methods for graphical models. *Machine Learn.* 37(2):183–233.
- Lee TY, Bradlow ET (2011) Automated marketing research using online customer reviews. *J. Marketing Res.* 48(5):881–894.
- Liu J, Hill S (2021) Frontiers: Moment marketing: Measuring dynamics in cross-channel ad effectiveness. *Marketing Sci.* Forthcoming.
- Liu J, Toubia O (2018) A semantic approach for estimating consumer content preferences from online search queries. *Marketing Sci.* 37(6):930–952.
- Liu J, Toubia O (2020) Search query formation by strategic consumers. *Quant. Marketing Econom.* 18:155–194.
- Liu T-Y (2009) Learning to rank for information retrieval. *Foundations Trends Inform. Retrieval* 3(3):225–331.
- Narayanan S, Kalyanam K (2015) Position effects in search advertising and their moderators: A regression discontinuity approach. *Marketing Sci.* 34(3):309–472.
- Netzer O, Feldman R, Goldenberg J, Fresko M (2012) Mine your own business: Market-structure surveillance through text mining. *Marketing Sci.* 31(3):521–543.
- Radinsky K, Svore KM, Dumais ST, Shokouhi M, Teevan J, Bocharov A, Horvitz E (2013) Behavioral dynamics on the web: Learning, modeling, and prediction. *ACM Trans. Inform. Systems* 31(3):16.
- Salakhutdinov R, Mnih A (2008) Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. *Proc. 25th Internat. Conf. Machine Learn.* (ACM, New York), 880–887.
- Sievert C, Shirley KE (2014) LDAvis: A method for visualizing and interpreting topics. *Proc. Workshop Interactive Language Learn. Visualization Interfaces* (Association for Computational Linguistics, Shroudsburg, PA) June 2014:63–70. Sponsor: Idibon.
- Statista (2017a) Paid search advertising expenditure worldwide from 2015 to 2017. Accessed January 2018, <https://www.statista.com/statistics/267056/paid-search-advertising-expenditure-worldwide/>.
- Statista (2017b) Statistics and facts about the television industry. <https://www.statista.com/topics/977/television/>.
- Tirunillai S, Tellis GJ (2014) Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation. *J. Marketing Res.* 51(4):463–479.
- Varian HR (2007) Position auctions. *Internat. J. Indust. Organ.* 25(6):1163–1178.
- Yang S, Ghose A (2010) Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence? *Marketing Sci.* 29(4):602–623.