# The Effects of Query Bursts on Web Search

Ilija Subašić
*Katholieke Universiteit Leuven*
*Leuven-Heverlee, Belgium*
*ilija.subasic@cs.kuleuven.be*

Carlos Castillo
*Yahoo! Research*
*Barcelona, Catalonia Spain*
*chato@yahoo-inc.com*

*Abstract*—**A query burst is a period of heightened interest of users on a topic which yields a higher frequency of the search queries related to it. In this paper we examine the behavior of search engine users during a query burst, compared to before and after this period. The purpose of this study is to get insights about how search engines and content providers should respond to a query burst.**

**We analyze one year of web-search logs, looking at query bursts from two perspectives. First, we adopt the user's perspective describing changes in user's effort and interest while searching. Second, we look at the burst from the general content providers' view, answering the question of under which conditions a content provider should "ride" a wave of increased interest to obtain a significant share of clicks.**

## I. INTRODUCTION

Users express a heightened interest in queries related to current events, leading to sharp increases in their frequency in web-search query logs. For instance, on October 18, 2008, after being parodied several times in the TV show Saturday Night Live, U.S. politician Sarah Palin appeared in the show and met her impersonator. This led to a $22\times$ increase in the frequency of the query "*snl sarah palin*" compared to two days before the event. This is referred to as a query burst [1].

In our research, we use an existing method to detect query bursts, and focus in characterizing their effects on web search. We look at query bursts from the users' perspective, trying to uncover how does the higher interest in a query change their behavior. Especially, we are interested in what happens before and after a query burst. We also look at query bursts from the content providers' perspective.

From an economics point of view, higher attention on a topic, quantified as query frequency, can be regarded as an increase in the demand for an informational good. The supply are the documents that are relevant for the topic. An increase in the demand generates an increase in the "price" users pay for accessing the information (quantified as the effort they spend searching), which is matched later by an increase in the supply of the informational good. If the demand drops afterwards, on the supply side documents are not created at the same pace as during the period of high activity. We are interested in how this process affects content providers in general, and in particular how it changes the share of attention that each content provider receives.

**Objectives.** We wish to analyze query bursts in order to understand what happens to the web during them. The hypothesis is that bursts have some impact, temporarily changing the behaviour of some users, inducing changes on web search rankings, and prompting content providers to generate new content. Our goal is to quantify these changes, see if they match the intuitions, and describe their effects on the different stake-holders in the web search process. Taking into account different perspectives we formulate the following research objectives: (i) characterize and describe different types of query bursts, (ii) measure changes in user's behavior, particularly effort and attention, and (iii) describe how content providers can take advantage of a query burst.

**Contributions.** This study contributes to the above objectives by observing that:
- Query bursts are not equal among them, but can be grouped in classes having distinctive statistical properties.
- During a query burst, not only query frequency, but per-query user effort is higher according to several metrics.
- During a query burst, clicks on search results tend to be more concentrated at the top.
- After a query burst, the distribution of clicks into search results changes substantially.
- Publishing early represents a clear advantage for content providers, and for some queries this advantage is unsurmountable. For other queries, a late-comer has an opportunity of obtaining a non-trivial share of the users' attention.

**Roadmap.** The next section describes previous work related to ours. Section III defines the concepts we use. Section IV describes in detail our experimental setting, sampling methods and metrics. Section V presents a classification of query bursts based on evidence from search logs. Section VI describes web search before, during and after a query burst. Section VII models changes in click share. Finally, Section VIII presents our conclusions.

## II. PREVIOUS WORK

Query-log analysis is a research topic that has received substantial attention in the last few years, with even entire venues devoted to the topic, such as the *Web Search Click Data* and the *Query Log Analysis* workshops. Since the early

studies of query logs presented in [2], [3], [4], the field has branched out into several areas, and our coverage of them in this brief section is by no means complete.

**Query categories.** The analysis of an hourly time series in [5] and a long-term time series in [6] show distinct properties in the frequency profile for queries belonging to different topical categories. Conversely, the authors of [7] study if the different frequency profiles can be used to improve query classification. In [8], [9] instead of topical categories authors look for differences between high-frequency and low-frequency queries. In the present study, we do not categorize general queries but only bursty ones, and our categories are based on multiple factors which are neither topics nor overall frequencies.

**Temporal query analysis.** Adar et al. [10] compare time series of queries from different sources. Their results are a description of different classes of temporal correlation and a visual tool for summarizing them. Previously, using correlation between query frequency time series [11] uncovered semantic similarity between time-aligned series. Time-based query similarity discovery is also the topic of [12], where clustering of a bipartite graph of queries and pages is used. Reference [13] presents a method to uncover possible causal relationships between queries. In contrast to previous works, our paper focuses on differences in user behavior before and after a certain disruptive event, and compares it to user behavior on randomly-chosen queries and on queries that are stable over time.

Our research over a 1-year period can be considered long-term with respect to a majority of works on query-log mining, with few exceptions e.g. [14].

**Query bursts.** Current applications of query-burst detection include the detection of real-world events [15], tracking the spread of diseases as flu [16] and in general tracking of memes/"buzz" in text corpora [1], [17], [18], [19]. In [20], query bursts are detected as outliers in the query frequency series, specifically as moments at which the query shows 1.5-2 standard deviations higher frequency than its average in previous periods. In [14], increases in normalized query frequency are used to discover query bursts. We adopt this method, which allows us to incorporate other constraints to query bursts detection. We wish to study the behaviour of bursty queries and therefore minimize the errors of the burst detection algorithm, increasing its precision at the expense of its recall. Our work goes beyond the detection of query bursts and explore their effects on users, web search, and content providers.

## III. PRELIMINARIES AND NOTATION

The concept of a "query burst" is quite broad and therefore there is no standard or widely-accepted test for detecting it. In this paper we are interested in identifying the query bursts precisely, and therefore look for clear outliers. We apply a bursty measure based on normalized lift in query probability, that has been successfully used for finding bursty queries in query logs [14] as well as in news documents [17]. We impose a large increase in frequency, and the property of having a single distinctive burst during the 1-year observation period. In practice, this turns out to be more stringent than the test shown in [20]: the query bursts we sample are very clear and would be detected as bursts by any reasonable test. Some examples are shown in Figure 1(a).

**Query burstiness.** Let $\mathcal{Q}$ be the set of all queries. Let $\mathcal{T}$ be the set of observation periods $\mathcal{T} = \{t_0, t_1, \ldots, t_{|\mathcal{T}|-1}\}$, in which each period represents an interval of time, for concreteness in this study each $t \in \mathcal{T}$ corresponds to one day. Let $f : (\mathcal{Q} \times \mathcal{T}) \to \mathbb{N}$ be such that $f(q,t)$ is the number of occurrences of query $q$ in the period $t$.

For each query $q$ and period $t$ we derive a BURST INTENSITY index $b(q,t)$ which tells us how "bursty" this query is in that period by measuring its relative increase in frequency compared to the past. This is obtained by computing

$$b(q,t) = \frac{f(q,t)/\sum_{q'\in\mathcal{Q}} f(q',t)}{\sum_{u\leq t} f(q,u)/\sum_{q'\in\mathcal{Q}}\sum_{u\leq t} f(q',u)} \ .$$

For a given $\beta$, whenever $b(q,t) \geq \beta \sum_{u\in\mathcal{T}} b(q,u)/|\mathcal{T}|$, we say that the query $q$ is going through a query burst at time $t$. If a query has no bursty period, we say the query is *non-bursty*. If the query has bursty periods that are not contiguous, we say the query is *bursty during multiple episodes*. If all the periods in which the query is bursty are contiguous, we say the query is *bursty during a single episode*. For each query that is bursty during a single episode, let $E_q = \{s_q, s_q+1, s_q+2, \ldots, s_q+d_q-1\}$ be the set of consecutive periods in $\mathcal{T}$ were the query is undergoing a query burst. We name $s_q$ the *start* of the episode, and $d_q$ the *duration* of the episode. In our experiments we select only bursty queries having a minimum episode duration $d_q \geq \delta$.

In the following, we refer to a sample of such queries as the BURSTY queries. We also built a sample of queries having a very small variation of $b(q,t)$ in the observed series. In the following we refer to this sample as the STABLE. These samples represent extremes; most queries are neither STABLE nor BURSTY, so for some experiments we introduce a sample of RANDOM queries chosen uniformly at random.

**Pre-episode, episode, and post-episode.** We also obtain time intervals before and after the episode for comparison, and refer to them as *pre-episode* and *post-episode*. These time intervals are obtained in such a way that they (i) are not too close to the episode, and (ii) comprise a number of occurrences of a query that is at most the occurrences of the query in the episode.

Formally, for a query having an episode starting at $s_q$ and having a duration $d_q$, the pre-episode ends at time $s_q - d_q$,
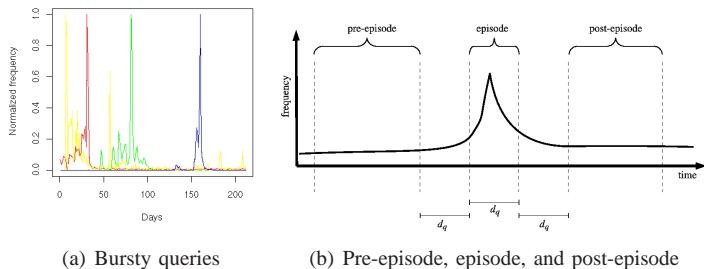
(a) Bursty queries     (b) Pre-episode, episode, and post-episode

Figure 1. Example bursty queries and depiction of a query burst.

and starts at a time $\mathrm{pre}(q)$ such that

$$\sum_{\mathrm{pre}(q)\leq t\leq s_q-d_q} f(q,t) \approx \sum_{t\in E_q} f(q,t)$$

in which the approximation is due to the time granularity of one day, so we approximate $\mathrm{pre}(q)$ to the nearest possible whole day. If there are not enough query occurrences before the episode, we set $\mathrm{pre}(q) = t_0$. We do the same for the post-episode period, starting at $s_q + 2d_q$ and ending at $\mathrm{post}(q)$ so that the total frequency during the post-episode period is at most the total frequency during the episode. If there are not enough query occurrences, we set $\mathrm{post}(q) = t_{|\mathcal{T}|-1}$.

Figure 1(b) depicts the relationship between pre-episode, episode, and post-episode.

**Pseudo-episodes:** For some experiments we want to study if a phenomenon is related to the bursty nature of the query or not. In the case of STABLE and RANDOM queries, we create *pseudo-episodes* that have the same query volume as the episodes of BURSTY queries, but usually a longer duration.

Specifically, for each of the queries in these samples, we select a starting date uniformly at random (leaving the first-3 and the last-3 months out), then pick the volume of queries in the *pseudo-episode* according to the distribution of query volume in the episodes of the BURSTY sample. The pre- and post-episode periods are created in the same manner as for the bursty queries.

## IV. EXPERIMENTAL FRAMEWORK

We processed a query log from Yahoo! to obtain one year of searches originated in the US. From this log we sampled three subsets, the BURSTY queries subset, the STABLE queries subset and the RANDOM query subset.

**Sampling.** Given the amount of data to process, we did an iterative process. We started by sampling 1,400 "torso" queries (having frequencies that were neither too low nor too high), and their sessions. After this, we continued by rounds deepening (sampling more sessions) and narrowing (sampling less queries) our sample. The process was completed with a full sample of all the user sessions during 13 months containing 190 queries that are *bursty during a single period*. In our experiments we set $\beta \geq 3.5$, meaning that

the $b(q,t)$ index must be 3.5 times larger than the average. We also set $\delta \geq 3$, meaning that the duration of the single episode must be of 3 days or more.

For the STABLE set, we set the maximum standard deviation of $b$ to 0.5 during the entire year, obtaining a set of 768 stable queries candidates, and sub-sampled a set of 200 queries according to the distribution of frequencies of the BURSTY sample. In other words, for any given frequency, both the BURSTY and the STABLE sets have approximately the same number of queries.

To select the RANDOM queries we first binned the bursty queries based on their frequency during the episode. Then from each bin we randomly selected queries having a 1-year frequency at most three times larger to ensure that pseudo-episodes have complete pre-episodes and post-episodes periods. Using this process we created a sample of 340 queries form the initial 1,400 "torso" queries sample.

**Metrics.** Our study of bursty queries is based on the analysis of a set of metrics that covers different aspects of the search queries.

*1. Activity/effort metrics:* The first group of metrics captures users' effort in finding the information they searched for. Most of these metrics are session-level, in which by *session* we mean groups of related queries, known in the literature as query chains [21] or search missions [22]. To obtain them, we segmented the activity of users into logically-coherent groups of queries, using the method in [23].

For a given query $q$, these metrics include:
- SESSION DURATION: average duration in seconds of sessions containing $q$.
- DWELL-TIME: average time in seconds from an occurrence of $q$ to the next query event, limited to 30 minutes.
- QUERIES/SESS.: average number of queries in sessions containing $q$.
- CLICKS/SESS.: average number of clicks in sessions containing $q$.
- EVENTS/SESS.: average number of events per session, including queries, clicks, and "previous-page/next-page" events.
- CLICKS/QUERY: number of clicks on average after a query $q$ and before any other query in the session.
- NON-CLICKS %: fraction of issued queries not followed by a click.
- ASSISTANCE %: fraction of query reformulations that were the result of a search suggestion. Most search engines display for some queries a few suggested queries, usually with a label such as "also try" or "related searches". This variable measures how often, when doing a reformulation, users click on one of these suggestions instead of typing by themselves a new query.
- USERS/QUERY: number of distinct users issuing $q$ divided by number of occurrences of $q$. This is in $[0,1]$ and values

close to 0 indicate that a small group of users is repeatedly issuing the same query.

*2. Attention metrics:* The second group of metrics describes how concentrated or disperse are users' clicks. For a particular period (episode or pre/post-episode) and a specific query, we sort the URLs clicked for that query during the period in decreasing order of click probability. In the following, by *top URL(s)* for a period we always mean the most clicked ones. This usually, but not always, matches the ordering in which URLs are shown to users, because of positional bias [24]. Our metrics include:

- DISTINCT URLS: number of distinct documents clicked.
- TOP-1 SHARE: probability of a click on the URL having the highest share of clicks.
- URLS 90%: minimum number of documents required to cover 90% of users' clicks.
- RANK-CLICK DROP: steepness of rank-click frequency curve, measured by the exponent resulting of fitting a power-law to the curve.
- CLICK ENTROPY: entropy of the distribution of clicks, as used in, e.g. [25], [26]

*3. Comparative metrics:* The third group of metrics compares different periods of time (e.g.: pre-episode and post-episode), focusing on changes in their click probability distributions. The goal of these metrics is to discover what is the impact of the query burst on the share of users' attention received by different pages.

- CLICK DIVERGENCE: KL-divergence of click distributions. For a query $q$, a set of URLs $D$, and a time period $t$, let $P(d|q,t)$ be the probability of having a click on page $d \in D$ for query $q$ during period $t$. Then, the KL-divergence between two periods $t_1$ and $t_2$ is defined as: $D_q\langle t_1||t_2\rangle = \sum_{d \in D} P(d|q,t_1) \times \log{(P(d|q,t_1)/P(d|q,t_2))}$.
- TOP-1 CHANGE: difference in the probability of the URL with the highest probability in the first period with respect to the second period.
- TOP-N OVERLAP: overlap of URLs sorted by click probability, at position $n = 1$ and $n = 5$, between the two periods.

We also considered variations in the activity/effort and attention metrics, e.g.: differences in DISTINCT URLS.

*4. Global metrics:*

- PEAK BUILD-UP RATIO: for a URL $u$, this is the difference between the episode peak, and the first date in which $u$ is clicked in the search result list. This is normalized using the difference between the episode peak and the start of the dataset, as we explain in Section VII. For instance, a value of 1 indicates the URL has existed since the beginning of the observation period, and a 0 indicates it was created the day of the peak of the query burst.
- BURST INTENSITY: the $b$ index described in Section III.

## V. CLASSES OF QUERY BURSTS

Differences of topics covered by bursty queries suggest that the nature of underlying events which caused the bursts and the way they develop are also different. We wish to discover the different patterns of query bursts based on user behaviour during them. Therefore, the first application of the metrics we described in Section IV is to characterize different types of query bursts. Since there is no ground truth for this type of classification, we choose to discover different types of bursts using an unsupervised approach. For this we apply k-means clustering algorithm using all the metrics extracted as input features. Our main goal is a descriptive analysis of bursty queries with the goal of discovering features that point to different classes of bursty queries, rather than a predictive one predicting bursty queries or classes they belong to.

We experimented varying the number of clusters from 2 to 30 and found no clear evidence of an intrinsic number of clusters in the data (e.g.: looking at sum of distance square from clusters centroids, we observed no steep drop when increasing the number of clusters). After inspecting the clusterings, we decided to use the one with 3 clusters. A different number of clusters could have been used, and we decided this particular clustering because we noticed that it uncovered groups with distinct features and an easy-to-grasp interpretation, and because it was also useful in practice for the predictive task of Section VII.

**Type A: bursts that fade out completely afterwards.** These queries are not very frequent during the pre-episode, and fade away quickly in the post-episode. They have a high divergence (high CLICK DIVERGENCE) between the pre- and post-episode, meaning that the episode changes completely the search results for the query. There is also no strong authoritative URL (low TOP-1 SHARE, high CLICK ENTROPY), which partially explains why click share is so strongly affected by the episode.

This cluster contains many queries related to entertainment, some examples are: *kate wiliams, superbowl commercials 09, snl sarah palin, jett travolta*. Typical behaviour of this type can be represented by the query *snl sarah palin*. The mentioned TV show caused a huge increase of the queries' frequency, and created a new, previously non-existing, topic without an authoritative source. These are "buzz" topics that after an initial hype quickly lose the interest of the users.

**Type B: bursts that create new topics.** These queries are also not very frequent during the pre-episode, but contrary to Type A, they maintain some presence in the post-episode. They have a less dominant top URL (medium TOP-1 SHARE) and less click concentration (medium CLICK ENTROPY).

This cluster contains many queries related to new scientific/technical developments and events that have long-

term effects, for instance: *air car, kawasaki disease, 2008 olympics, joe biden, obama mccain polls*. For example, the information on *2008 olympics* is present long before the games commence, but it is the start of the games that triggers the increased user interest in the topic, and changes the click distribution to, in this case, sporting events result pages.

**Type C: bursts on existing topics.** These queries appear both in the pre-episode and in the post-episode with non-negligible frequency. They have an authoritative top result with a high click share (high TOP-1 SHARE) and a low CLICK ENTROPY, so the users' attention is concentrated. For these queries, the episode does not change the distribution of clicks, reflected by the fact that the CLICK DIVERGENCE is low.

This cluster contains many queries related to topics that are searched during the entire year, but for which a real-world event triggers heightened user interest. Examples: *irs refund status, teen choice awards, national hurricane center, saturday night live*. For example, the burst of *saturday night live* is caused by the same previously discussed TV appearance of U.S. politician Sarah Palin, but the query itself is present before that particular show and its burst does not have long-lasting effects on the search results for the query.

**Remark.** This classification of query bursts matches the classes of bursts predicted by the model of Crane and Sornette [27] using completely different methods. Type A corresponds to exogenous sub-critical, expected in cases of external events that do not propagate well virally. Type B corresponds to exogenous critical, expected in cases of external events that are highly viral. Type C corresponds to endogenous critical, expected in cases of internally-motivated messages that are highly viral.

## VI. BEFORE, DURING AND AFTER A QUERY BURST

If query burst represent the heightened interest of users in query-related topics, this should be reflected not only in the query frequency, but in the "cost" users will accept to find the content they seek. In search process this cost can be expressed as effort users are willing to put in, and the concentration on a certain parts of result set. To confirm this we look at specific sets of metrics, studying them during the *pre-episode*, *episode*, and *post-episode* periods defined as in Section III. We study changes in multiple query attributes during the query burst, as detailed in the rest of this section. To the test statistical significance of changes on a metric, we used either t-test or KolmogorovSmirnov test, depending on the distribution of the metric. Our findings can be summarized as: (i) not only frequency, but also per-user effort/activity is higher during the query burst. and (ii) users' clicks are more concentrated during the query bursts.

**1. User effort/activity is higher during query bursts.**

Table I shows an increase in several metrics of activity/effort for bursty queries during their episode compared

### Table I
AVERAGE OF ACTIVITY/EFFORT METRICS FROM SECTION VI IN PRE-EPISODE, EPISODE, AND POST-EPISODE; AND STABLE QUERIES. STATISTICALLY SIGNIFICANT DIFFERENCES WITH EPISODE: $p < .01$ (***), $p < .05$ (**), $p < 0.10$(*)

| Metric | Pre- | Episode | Post- | Stable |
|---|---|---|---|---|
| SESSION DURATION | 1768.6 | 1886.00 | 1624.10 | 2238.1** |
| DWELL-TIME | 175.13 | 178.00 | 157.80 | 216.7* |
| EVENTS/SESS. | 5.06*** | 7.64 | 4.57*** | 4.69*** |
| QUERIES/SESS. | 2.67*** | 3.19 | 2.28*** | 2.14*** |
| CLICKS/SESS. | 2.29*** | 3.73 | 1.96*** | 1.87*** |
| CLICKS/QUERY | 0.79 | 1.81 | 1.39 | 0.86*** |
| ASSISTANCE % | 11.90*** | 13.18 | 12.29*** | 4.69** |
| NON-CLICKS % | 35.97*** | 28.22 | 41.84*** | 22.25*** |
| USERS/QUERY | 1.47* | 1.65 | 1.47 | 2.87** |

### Table II
AVERAGE OF ATTENTION METRICS FROM SECTION VI. STATISTICALLY SIGNIFICANT DIFFERENCES WITH EPISODE: $p < .01$ (***), $p < .05$ (**), $p < .10$ (*)

| Metric | Pre- | Episode | Post- | Stable |
|---|---|---|---|---|
| TOP-1 SHARE | 0.52 | 0.56 | 0.52 | 0.71*** |
| RANK-CLICK DROP | 1.15*** | 1.01 | 1.10*** | 0.55*** |
| CLICK ENTROPY | 1.54** | 1.44 | 1.61*** | 0.96*** |
| URLs 90% | 5.12 | 4.40 | 5.46** | 4.69 |
| DISTINCT URLs | 32.95 | 35.57 | 41.03* | 59.17*** |

to their pre-episode and post-episode. During the episode, sessions are not significantly longer in duration, but contain more queries, more clicks, and more events in general; also more individual sessions have clicks.

Bursts of query activity are driven mostly by an increase in the number of users issuing the query, given that the ratio USERS/QUERY does not change significantly. The fact that on average users click on search assistance more often during the episode, may indicate less familiarity with the topic being queried; the comparison with the stable queries also points in that direction.

Compared to their pre-episode and post-episode, bursty queries *during their burst generate more "intense" search sessions*. This increase may be due to a number of causes, including increased interest and increased difficulty in locating information. Given that most episodes tend to last only a few days, the effect of the episode in effort and activity could be attributed more to increased user interest.

A possible explanation, whose empirical analysis goes beyond the scope of this paper, is that users who participate in a query burst become "activated" after the signals they receive go beyond an activation threshold (see e.g. [28]). In other words, users who query about a topic for the first time, must be sufficiently interested in the topic to query about it, and thus are willing to spend more effort searching.

*Comparison with stable queries:* in terms of activity/effort, stable queries are part of longer sessions with less events, hence longer dwell times. Stable queries also have much less use of search assistance.

**2. Clicks are more concentrated during episodes.**

Table II shows that clicks tend to be more concentrated during the query burst than in the pre-episode and post-

episode periods. The share of clicks of the single top URL does not change significantly, but click probabilities on the top clicked URLs are higher, as evidenced by a more steep rank-click drop and a lower entropy.

If we observe that clicks are more concentrated, and that sessions during query bursts have more clicks but do not last significantly longer, we must conclude that during the burst less results are examined (i.e.: users do not read the usual number of titles and snippets). If many users that arrive during the burst are unfamiliar with the topic, but want to quickly get informed, this is exactly what we would expect.

In the post-episode, there is an increase in the number of distinct URLs, and the number of documents required to cover 90% of the clicks. This indicates that new relevant documents are created during and after the query burst.

*Comparison with stable queries*: it is clear that stable queries have clicks that are even more concentrated at the top that in the case of bursty queries, according to all metrics we examined.

## VII. SEARCH RESULTS AND CLICK SHARE

Sections V and VI showed the differences in types of bursts, and users behaviour during the burst. In this section we turn to the search results, and investigate the effect of the query burst on the distribution of clicks on search results, referred in the following simply as the "click distribution".

Basically, we aim to discover if the query burst presents an opportunity for publishing a web page about the topic of the query burst. We expect that existing documents will have the largest share of clicks, but that perhaps new documents can also capture some clicks. Concretely, we investigate the following questions: (i) how much is the click distribution changed by the query burst?; (ii) is it necessary to have a page that existed before the burst to have a large share of clicks?; and (iii) is it possible to predict the share of new documents during the burst?

**Changes in click share:** we measure the effect of the episode in the click distribution using the CLICK DIVERGENCE measure defined previously. We compared the click distributions of pre-episode, episode, and post-episode for the BURSTY sample, and use *pseudo-episodes* for the RANDOM and STABLE samples.

The results shown in Figure 2(a) show clearly that the *click distribution changes significantly with a query burst*. The click distribution of BURSTY queries changes on average about $3\times$ and $6\times$ more than for RANDOM and STABLE queries respectively. If we focus on the top-5 results only, as in Figure 2(b), we see that the changes are smaller but the gap between BURSTY queries and the rest is even larger. The largest difference between the click distribution between pre- and post-episode is the effect of both the time difference between them and the shift of interest during the burst.



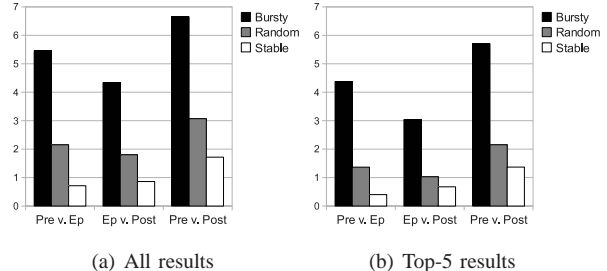(a) All results       (b) Top-5 results

Figure 2. Change in click distributions for BURSTY, RANDOM, and STABLE queries, measured using KL-divergence.

These results indicate that query bursts are disruptive events, that they can change the competitive landscape for content providers on a topic, shuffling rankings significantly.

**Click share of late-comers:** when the frequency of a query increases, most content providers that already have pages on the topic will receive an increased number of visits and will thus benefit from the heightened user's interest. The experimental data confirms that *publishing early gives a significant advantage*.

How large is this advantage? To quantify it, we use the metric PEAK BUILD-UP RATIO of a URL $u$ in query $q$. It measures how soon the URL appears in the query log in comparison with the query burst peak. Let $t_{u,q}^{first}$ be the first time the URL $u$ is clicked for query $q$, and let $t_q^{peak}$ be the time of the peak of the query burst of $q$. Let $t_0$ be the beginning of the observation period, then this metric is

$$\text{PEAK BUILD-UP RATIO}(u,q) = \max\left\{\frac{t_q^{peak} - t_{u,q}^{first}}{t_q^{peak} - t_0}, 0\right\}$$

and a value close to 1 means the URL's first click was close to the beginning of the observation period, while a 0 indicates the URL's first click occurred on the day of the peak. The first click in the URL could be observed *after* the episode peak, but this is a rare case and for simplicity of the presentation we truncate those values to zero. In the following, we will refer to documents whose PEAK BUILD-UP RATIO is close to zero as late-comers or *new pages*.

Figure 3(a) indicates that 61% of the top-URLs have existed since the beginning, while only 16% of the top-URLs are new pages. This is consistent with findings of [19] which show that "good" content providers publish before the burst starts. When looking at the top-10 results, on average about 3 pages are new, while in the bottom-10 results, on average about 5 pages are new. In general, Figures 3(b), 3(c), and 3(d) show that publishing late, i.e.: having PEAK BUILD-UP RATIO close to zero, means a lower share of clicks during the episode.

Next, we consider the *share* of clicks the new pages will obtain. This information is presented in Table III, where we take the Top-1, Top-5, Top-10, and All of the *newly-created URLs* and look at their click share. In general, the
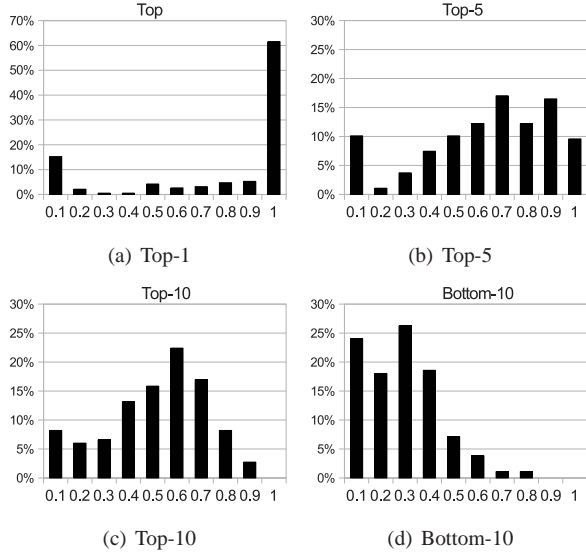
Figure 3. Average PEAK BUILD-UP RATIO for the (a) the top result, (b) the top-5 results, (c) the top-10 results, (d) the bottom-10 results.

Table III
CLICK SHARE OF THE NEW URLs AS A PERCENTAGE OF TOTAL CLICKS. TOP-$k$ INDICATES THE $k$ MOST CLICKED NEW URLs. "ALL" INDICATES ALL THE NEW URLs

| Query cluster | Top-1 | Top-5 | Top-10 | All |
|---|---|---|---|---|
| All queries | 3.1% | 5.5% | 8.9% | 27.5% |
| A: bursts that fade out completely afterwards | 37.8% | 41.1% | 20.2% | 52.1% |
| B: bursts that create new topics | 5.9% | 5.6% | 5.2% | 25.2% |
| C: bursts on existing topics | 2.5% | 3.5% | 4.2% | 9.8% |

new URLs obtain the minority of clicks during the episode, and this is distributed among many queries: even the Top-10 most clicked new pages considered together obtain less than 10% of the clicks.

Our findings from Section V suggest that the click share of at least the top-URL is different depending on the type of query burst. Therefore, Table III also includes per-cluster results, which show that *the best opportunities for late-comers are queries of type A* (bursts that fade out completely afterwards) where they can obtain roughly 50% of the users' clicks. On the other hand, queries of type C (bursts on existing topics) are basically dominated by documents published before the query burst, and the share of clicks obtained by late-comers is around 10%.

**Finding opportunities for late-comers.** From the content-providers' perspective, the question of finding which are the "waves" that should be ridden is a central one. The automatic method we present in this section can help them make this decision, but does not provide a perfect prediction. A system that were to help content providers in deciding what to write about, should be able of (a) identifying query bursts as they happen, and (b) predicting the expected benefit. Question (a)

was the subject of Section IV while (b) turns out to be more difficult.

As mentioned in the previous sections, the target of this prediction task is the click share of new pages. We first use a logistic regression model $M$ with the features from the pre-episode and episode described in the Section IV. Its performance, measured using the correlation coefficient between the predicted click share and the actual click share for a hold-out test set of queries (50% of sc Bursty set), are 0.42 when considering all results, and 0.69 when considering the top-10 documents.

The insights from Table III can be used to improve this prediction, given that the average share of newly published pages depends on the cluster to which the query belongs. Thus, we build a model $M'$ that first computes the probability of a query belonging to each cluster using a Naive Bayes classifier, and then includes these predictions in the logistic regression model. The correlation coefficient between the original and predicted value increases to 0.46 when considering all the results, and to 0.77 when considering the top-10 documents. Although the model for predicting burst types does not have high performance ($F-1$ measure $=0.72$), the usage of the cluster prediction in $M'$ improves the prediction of $M$. These results show that it is hard to predict the values for all the pages, while a fair performance can be obtained for the Top-10 results.

## VIII. CONCLUSIONS

Query bursts are observed in a search engine query log whenever there is increased interest in a certain topic. Our research over a 1-year-long query log from a major search engine uncovered different types of query bursts. When we grouped query bursts into three classes, we found (A) bursts that fade out completely afterwards, (B) bursts that create new topics, and (C) bursts on existing topics.

The analysis of several metrics provides fairly intuitive results backed up by empirical evidence. During query bursts users do not spend much more time searching than usually, but clearly invest more effort in terms of number of queries and clicks per session. Also, their clicks are more concentrated on a smaller group of search results. After the query burst, the distribution of clicks into search results for the query is substantially different from that before the query burst, as new relevant pages are created and rankings change.

**Content providers** that intend to capture users' attention on emerging topics should attempt to publish early. If not, they should target query bursts on emerging topics that did not exist before (types A and B). Writing about a previously-existing topic (type C) during a query burst is unlikely to yield a substantial share of clicks. We developed predictive models, providing a method to obtain a rough estimate of the share of clicks that can be obtained from a query burst by content providers that did not publish early about a topic. Although content providers do not have a direct access to

query-logs, search engines could provide a sort of a topic "recommendation" based on the interest of the users visiting a specific URL or a set of URLs.

**Search engines** should, according to our findings, treat queries undergoing query bursts differently. For instance, search suggestions are clicked much more frequently for these queries. In general, search engines should introduce changes to support the needs of users entering bursty queries.

We consider this work as a part of a broader effort, which is to provide the right signals about users' needs to content providers. Search engines should help to detect scarcity of information on certain topics so that content providers can supply this information. This involves creating models that also take into account content providers' features such as topic, influence and authority, and that are able to detect users' unsatisfied needs for information in certain areas. This is the goal towards which we plan to orient our research in the future.

Key references: [10], [14]

## REFERENCES

[1] J. Kleinberg, "Bursty and hierarchical structure in streams," in *Proc. KDD*. Edmonton, Canada: ACM, 2002, pp. 91–101.

[2] B. J. Jansen, A. Spink, and T. Saracevic, "Real life, real users, and real needs: a study and analysis of user queries on the web," *Inf. Proc. & Mgmt.*, vol. 36, no. 2, pp. 207–227, March 2000.

[3] T. Lau and E. Horvitz, "Patterns of search: analyzing and modeling web query refinement," in *Proc. UM*. Banff, Canada: Springer, 1999, pp. 119–128.

[4] C. Silverstein, H. Marais, M. Henzinger, and M. Moricz, "Analysis of a very large web search engine query log," *SIGIR Forum*, vol. 33, no. 1, pp. 6–12, 1999.

[5] S. M. Beitzel, E. C. Jensen, A. Chowdhury, D. Grossman, and O. Frieder, "Hourly analysis of a very large topically categorized web query log," in *Proc. SIGIR*. Sheffield, UK: ACM Press, 2004, pp. 321–328.

[6] S. M. Beitzel, E. C. Jensen, A. Chowdhury, O. Frieder, and D. Grossman, "Temporal analysis of a very large topically categorized web query log," *J. Am. Soc. Inf. Sci. Technol.*, vol. 58, no. 2, pp. 166–178, 2007.

[7] S. Asur and G. Buehrer, "Temporal analysis of web search query-click data," in *Proc. SNA-KDD*. Paris, France: ACM Press, 2009.

[8] D. Downey, S. Dumais, and E. Horvitz, "Heads and tails: studies of web search with common and rare queries," in *Proc. SIGIR*. Amsterdam, The Netherlands: ACM, 2007, pp. 847–848.

[9] D. Downey, S. Dumais, D. Liebling, and E. Horvitz, "Understanding the relationship between searchers' queries and information goals," in *Proc. CIKM*. Napa Valley, CA, USA: ACM, 2008, pp. 449–458.

[10] E. Adar, D. S. Weld, B. N. Bershad, and S. S. Gribble, "Why we search: visualizing and predicting user behavior," in *Proc. WWW*. Banff, Canada: ACM Press, 2007, pp. 161–170.

[11] S. Chien and N. Immorlica, "Semantic similarity between search engine queries using temporal correlation," in *Proc. WWW*. Chiba, Japan: ACM Press, 2005, pp. 2–11.

[12] Q. Zhao, S. C. H. Hoi, T.-Y. Liu, S. S. Bhowmick, M. R. Lyu, and W.-Y. Ma, "Time-dependent semantic similarity measure of queries using historical click-through data," in *Proc. WWW*. Edinburgh, Scotland: ACM, 2006, pp. 543–552.

[13] Y. Sun, K. Xie, N. Liu, S. Yan, B. Zhang, and Z. Chen, "Causal relation of queries from temporal logs," in *Proc. WWW*. Banff, Canada: ACM Press, 2007, pp. 1141–1142.

[14] M. Richardson, "Learning about the world through long-term query logs," *ACM Trans. Web*, vol. 2, no. 4, pp. 1–27, 2008.

[15] Q. Zhao, T.-Y. Liu, S. S. Bhowmick, and W.-Y. Ma, "Event detection from evolution of click-through data," in *Proc. KDD*. Philadelphia, PA, USA: ACM, 2006, pp. 484–493.

[16] Google Inc., "Google Flu Trends," http://www.google.org/flutrends/, 2009.

[17] I. Subašić and B. Berendt, "Web mining for understanding stories through graph visualisation," in *Proc. ICDM*. Pisa, Italy: IEEE Press, 2008.

[18] R. Kumar, J. Novak, P. Raghavan, and A. Tomkins, "On the bursty evolution of blogspace," in *Proc. WWW*. Budapest, Hungary: ACM Press, 2003, pp. 568–576.

[19] J. Leskovec, L. Backstrom, and J. Kleinberg, "Meme-tracking and the dynamics of the news cycle," in *Proc. KDD*. Paris, France: ACM, 2009, pp. 497–506.

[20] M. Vlachos, C. Meek, Z. Vagena, and D. Gunopulos, "Identifying similarities, periodicities and bursts for online search queries," in *Proc. SIGMOD*. Paris, France: ACM Press, 2004, pp. 131–142.

[21] F. Radlinski and T. Joachims, "Query chains: learning to rank from implicit feedback," in *Proc. KDD 2005*. Chicago, IL, USA: ACM Press, 2005, pp. 239–248.

[22] R. Jones and K. L. Klinkner, "Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs," in *Proc. CIKM 2008*. Napa Valley, USA: ACM Press, Oct 2008.

[23] P. Boldi, F. Bonchi, C. Castillo, D. Donato, A. Gionis, and S. Vigna, "The query-flow graph: model and applications," in *Proc. CIKM*. Napa Valley, USA: ACM, 2008, pp. 609–618.

[24] N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey, "An experimental comparison of click position-bias models," in *Proc. WSDM*. Stanford, CA, USA: ACM, 2008, pp. 87–94.

[25] J. Teevan, S. T. Dumais, and D. J. Liebling, "To personalize or not to personalize: modeling queries with variation in user intent," in *In Proc. of SIGIR '08*. New York, NY, USA: ACM, 2008, pp. 163–170.

[26] Q. Mei and K. W. Church, "Entropy of search logs: how hard is search? with personalization? with backoff?" in *Proc. WSDM*, Stanford, CA, USA, 2008, pp. 45–54.

[27] R. Crane and D. Sornette, "Robust dynamic classes revealed by measuring the response function of a social system," *In Proc. of the NAS*, vol. 105, no. 41, pp. 15 649–15 653, October 2008.

[28] M. Granovetter, "Threshold models of collective behavior," *The American Journal of Sociology*, vol. 83, no. 6, pp. 1420–1443, 1978.