Influences of a Shocking News Event on Web Browsing

Danai Koutra*
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA
danai@cs.cmu.com

Paul Bennett
Microsoft Research
One Microsoft Way
Redmond, WA
pauben@microsoft.com

Eric Horvitz
Microsoft Research
One Microsoft Way
Redmond, WA
horvitz@microsoft.com

ABSTRACT

It has been suggested that online search and retrieval contributes to the intellectual isolation of users within their preexisting ideologies, where people's prior views are strengthened and alternative viewpoints are infrequently encountered. This so-called "filter bubble" phenomenon has been called out as especially detrimental when it comes to dialog among people on controversial, emotionally charged topics, such as the labeling of genetically modified food, the right to bear arms, and online privacy. We seek to identify and study temporal information-seeking behavior and access to alternative versus reinforcing viewpoints following shocking, emotional, and large-scale news events. We choose for a case study to analyze search and browsing on gun control/rights, a strongly polarizing topic for both citizens and leaders of the United States. We study the period of time preceding and following a mass shooting to understand how its occurrence, follow-on discussions, and debate may have been linked to changes in the patterns of searching and browsing. For that purpose, we represent the web search and browsing behavior of the users as a Markov chain and study the transitions of the users between webpages of various views by using mobility indices.

1. INTRODUCTION

How do people transition between webpages on polarizing topics? Do shocking news events burst their ideological bubbles and make them more likely to seek information on opposing views? Understanding the web dynamics and change in the user browsing behavior over time are the key points we investigate in this work.

With advances in personalization methods, search engines and recommendation systems increasingly adjust results to users' preferences, as inferred from their past searches and choices. In addition, users often input biased queries [9], which reflect their own positions, while personalized results have the potential to just reinforce these opinions, acting as echo chambers. As a result, according to several recent studies [6, 9], the users remain within their informational bubbles. The phenomenon is sometimes referred to as the "filter bubble" effect, where users get exposed only to opinions that align with their current views. This effect, where the "web world" does not reflect the richness of views in the "real world" may be exacerbated for polarizing topics. We take as polarizing or controversial topics those linked to opposing perspectives, such as abortion, gun control vs. rights, labeling of genetically modified food, and privacy in online services.

In order to understand the users' web transitioning behaviors on polarizing issues, we focus on a highly controversial topic in the US: gun control and rights. At one end of the spectrum, extreme gun rights supporters argue an interpretation of the 2nd Amendment to the US Constitution that would prohibit any regulation of firearms. On the other side of the spectrum, extreme gun control supporters advocate the total ban of any private citizen ownership of firearms. In addition to these two extreme opinions, there exist multiple variations which lie between them (*e.g.*, more background checks, ban of fully automatic firearms). For our study we use web browser toolbar logs from November and December 2012, and primarily consider two time periods: before and after the Sandy Hook Elementary School shootings (S.H.) in Newtown, Connecticut (December 14th), an event with broad news coverage and nationwide impact. The event clearly had considerable influence on information seeking about gun control related topics as signified by the increased user activity in the days following the event (see Fig. 1). Both the first big spike in the figure, which corresponds to visits to on-topic websites on the day of the shootings, and other important spikes have been annotated. The effect on the quantity of information seeking is indisputable; so our focus is *not* on the increase in user activity, *but* on whether (and how) the event changed the *type* of activity.

For our analysis, we use raw web browser toolbar visitation logs from a popular commercial web browser where users have given consent to logging all non-https URLs from URLs visited from search and those reached by direct entry or browsing. By processing the logs, we obtained a large-scale dataset of user interaction data that is relevant to the gun debate, constituting about 61 000 distinct users visiting 297 000 on-topic websites, which were labeled according to their stance. This large corpus enabled the exploration of the click trails of the users to understand how people transition among webpages of opposing views, and how news about the shootings influences such transitions.

2. RELATED WORK

Our work is related to research on political controversies, conjectures about the so-called filter bubble, and the temporal evolution of knowledge. In the area of *political controversies*, where the interest lies in methods for polarity detection and political leaning classi-

*Work done during an internship at Microsoft Research.

TAIA '14 Gold Coast, Australia

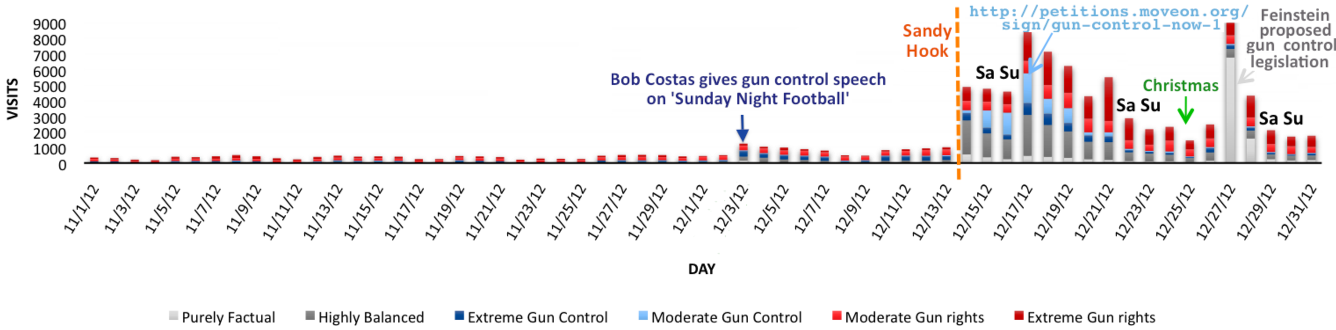

Figure 1: Number of visits to webpages that discuss gun control and rights issues over time (November-December 2012). The colors correspond to webpages categories: gray for factual and balanced pages; blue for pages supporting gun control; and red for pages supporting gun rights. Details about the categories and the labeling process are provided in the extended version of our paper [3].

fication, such as [1, 10]. Pariser [5] points out the existence of the *filter bubble*, which he defines as "this unique, personal universe of information created just for you by an array of personalizing filters", while others including [9, 6] propose ways to mitigate its effects and quantify it. Finally, *temporal evolution of knowledge* has been studied based on web logs in terms of changes in vocabulary, sites visited, and search strategies [8, 2].

This work. In contrast to most previous work, which considers primarily news outlets and blogs, and studies whether people access sources of different political categories to get informed [9], we put a particular *event* under the microscope and study how *that* affects the browsing behavior of the users on relevant topics. Thus, we do not engage in the common practice of characterizing websites as liberal and non-liberal, but define our own labels that are related to the stance of the pages on the particular topic of interest. Finally, our goal is to understand, around a shocking news event, the web dynamics, and specifically how people transition among content expressing different viewpoints, and how their behavior changes over time.

3. DATA

We use a proprietary dataset comprising anonymized web browser toolbar logs from November and December 2012 for more than 29 million users in the US-English market. The web logs include search and browsing behavior for the users, covering issued queries and visited, non-encrypted URLs. The selected users constitute a fair and broad sample of the US population. We consider primarily two time periods: before and after the Sandy Hook Elementary School shooting on December 14th. We note that we consider logs from a longer period of time before the event to develop a more robust estimate of users' habitual activity – a similar quantity of activity is observed in the period after the shooting because information seeking is more frequent after the event (Fig. 1). For the purposes of our study, we consider the URLs that are *on-topic*, i.e., websites that discuss gun control/rights issues. The technical details of the data extraction process are given in a longer technical report of this paper [3].

The resulting dataset, GC-DEBATE, consists of records `<user-id, session-id, URL, timestamp>` (Table 1). In the following sections, we refer to the intersection between the sets of users before and after the shootings as **common users**. Studying them enables us to directly compare changes in user behavior by controlling for the set of users.

To analyze the web browsing behaviors of the users, the webpages have to be labeled based on their stance on gun control/rights. Rather than focusing on alignment with a political party, we focus on the disposition of the content itself. Visits to a site that is predominantly affiliated with one party (*e.g.*, Democratic/Republican) or a particular pundit, does not by itself imply a lack of diversity in content; sites may contain content discussing a broad range of material. From the extracted web corpus, we label every page that is not off-topic or not accessible. In our analysis, we consider two sets of labels (Table 2):

Table 1: GC-DEBATE dataset. The last column holds the number of common users, URLs and domains between the two time periods.

	Before S.H.	After S.H.	Total	Overlap
Users	10,336	54,849	61,276	3,909
Sessions	20,656	70,138	90,794	N/A
Unique URLs	5,955	20,090	24,439	1,606
Unique Domains	277	501	613	165
Total Visits	118,749	177,975	296,724	N/A

- High-level labels, which classify the webpages to three broad categories, namely "Factual/Balanced", "Gun Control" and "Gun Rights".
- Expanded labels, which additionally identify the stance of the web content and includes symmetric and objective categories: "Purely Factual", "Extreme Gun Control", "Moderate Gun Control", "Highly Balanced", "Moderate Gun Rights", "Extreme Gun Rights".

More details on the labeling procedure are given in [3].

4. WEB TRANSITIONS

Given the search and browsing behavior of thousands of user on the topic of gun control and rights, we want to understand the way users navigate polarizing topics around a disruptive news event. Specifically, how does the current website's stance affect how people transition to the next webpage on polarizing political topics? What webpages are they more likely to visit after browsing a site supporting extreme gun control or rights? Although most of the users are not very diverse in terms of the stance of the domains they visit, many of them *do* transition between pages supporting opposing views. We seek to understand the most common transitions, as well as possible changes in the transitions due to the news on the Newtown shootings. By focusing on the influences of the stance of the current page on transitions to next pages, we obtain a micro-level view of information consumption patterns.

For each user we represent her browsing history as a Markov chain with either the high-level or expanded labels as states X_i

Table 2: Markov chain states for browsing history: Abbreviations for the high-level and expanded labels.

High-level States (3)		Expanded States (6)	
GC	Gun Control	**EC**	Extreme Gun Control
		MC	Moderate Gun Control
BF	Balanced/Factual	**HB**	Highly Balanced
		PF	Purely Factual
GR	Gun Rights	**MR**	Moderate Gun Rights
		ER	Extreme Gun Rights

(Table 2). Then, we describe the distribution of the transition probabilities by an n-state transition matrix $\mathbf{P}^n$, with elements $p_{ij} = Prob(X_{t+1} = j|X_t = i)$. We note that the row-wise sums are equal to 1, $\sum_j p_{ij} = 1$. To make sense of the underlying trends of this matrix, we employ mobility indices that have been widely used in economics and sociology for credit mobility and social status mobility among others.

4.1 Overall Transition Patterns

First, we consider the transitions of *all* users during November and December. There are $\sim 10K$ total transitions, corresponding to more than 7K users. We note that these are the users who visited at least two different domains, and, hence, we record for them at least one transition. The transitions are given in the form of a transition matrix in Tables 3 and 4.

Table 3: *All* users: 6-state transition matrix $\mathbf{P}^6$ for November-December.

	EC	MC	HB	PF	MR	ER
EC	32.65%	3.21%	24.20%	7.58%	11.66%	20.70%
MC	15.22%	2.17%	27.17%	8.70%	22.83%	23.91%
HB	10.53%	2.26%	28.07%	14.04%	16.42%	28.70%
PF	5.83%	3.79%	27.11%	10.50%	20.99%	31.78%
MR	6.28%	3.07%	18.83%	11.09%	21.90%	38.83%
ER	5.49%	1.55%	15.93%	9.36%	19.18%	48.49%

Table 4: *All* users: 3-state transition matrix $\mathbf{P}^3$ for November-December.

	GC	BF	GR
GC	28.42%	40.03%	31.55%
BF	12.00%	45.64%	42.36%
GR	5.55%	27.74%	66.71%

We first employ the so-called *Summary Mobility Indices*, which describe the direction of the mobility:

- Immobility Ratio: $IR = \sum_{i=1}^{n} p_{ii}/n$
- Moving Up: $MU = \sum_{i<j} p_{ij}/n$
- Moving Down: $MD = \sum_{i>j} p_{ij}/n$,

where n is the number of states. The immobility ratio represents the same-state transitions (higher for more *im*mobility), while the other two indices give the percent of transitions from one extreme to the other, *i.e.*, the MU index captures the transitions from extreme gun control towards extreme gun rights, and the MD index describes the reverse directionality (higher for more mobility).

The Summary Mobility Indices for all users during November and December are: (a) for the high-level states $IR = 0.4692$, $MU = 0.3798$, and $MD = 0.1510$, and (b) for the extended states $IR = 0.2486$, $MU = 0.4997$, and $MD = 0.2518$. Firstly, we observe that the overall system is characterized by mobility ($IR \ll 1$). Specifically, for the extended states, about 25% of the transitions are same state, and 50% of the transitions occur in the direction from extreme gun control towards extreme gun rights. From the transitions in the opposite direction, the most dominant transitions are towards the "middle" states: from factual to balanced webpages (27.11%), from extreme to moderate gun rights (19.18%), and from moderate gun rights to balanced pages (18.83%). All in all, the users mainly browse domains of the same stance or transition from gun control and balanced pages to websites supporting gun rights.

4.2 Sandy Hook: Change in Transition Patterns

As we seek to understand the effects of Sandy Hook to the opinions people are exposed, as previously, we restrict our analysis to the common users, and create two transition matrices, $\mathbf{P}_{before}$ and $\mathbf{P}_{after}$ for the expanded states shown in Tables 5 and 6 respectively, as well as the high-level states given in Table 7.

Table 5: *Common* users: Transition matrix $\mathbf{P}^6_{before}$ for the time period before Sandy Hook.

	EC	MC	HB	PF	MR	ER
EC	49.38%	0.00%	18.52%	3.70%	16.05%	12.35%
MC	11.76%	0.00%	23.53%	0.00%	23.53%	41.18%
HB	8.82%	1.68%	26.47%	8.40%	27.31%	27.31%
PF	2.90%	1.45%	30.43%	4.35%	33.33%	27.54%
MR	2.97%	1.30%	9.65%	6.86%	42.67%	36.55%
ER	1.82%	0.13%	10.29%	2.86%	20.70%	64.19%

Table 6: *Common* users: Transition matrix $\mathbf{P}^6_{after}$ for the time period after Sandy Hook.

	EC	MC	HB	PF	MR	ER
EC	22.03%	3.95%	26.55%	7.91%	12.99%	26.55%
MC	13.85%	0.00%	15.38%	13.85%	16.92%	40.00%
HB	9.01%	1.93%	21.89%	21.24%	15.02%	30.90%
PF	2.68%	3.68%	19.73%	8.36%	20.07%	45.48%
MR	3.97%	1.51%	15.69%	14.74%	18.15%	45.94%
ER	2.85%	0.95%	8.68%	12.47%	13.69%	61.36%

Table 7: *Common* users: 3-state transition matrices $\mathbf{P}^3_{before}$, $\mathbf{P}^3_{after}$.

	$\mathbf{P}^3_{before}$				$\mathbf{P}^3_{after}$		
	GC	BF	GR		GC	BF	GR
GC	42.86%	22.45%	34.69%	**GC**	22.73%	33.06%	44.21%
BF	9.12%	34.85%	56.03%	**BF**	9.15%	37.25%	53.59%
GR	2.91%	14.54%	82.56%	**GR**	4.24%	23.60%	72.16%

We start with the Summary Mobility Indices, as well as the eigenvalue-based indices [7, 4] that quantify the amount of mobility in the system. This category includes the eigenvalue M_E, second eigenvalue M_2, determinant M_D, and Singular Value Decomposition M_{SVD} indices. A value of 0 means to total immobility, and a value of 1 to perfect mobility.

Table 8: *Common* users: Summary Mobility and Eigenvalue-based Indices for $\mathbf{P}^3_{before}$ and $\mathbf{P}^3_{after}$. The top rows correspond to transitions between the 3 high-level states. Similarly, the bottom rows correspond to transitions between the 6 expanded states.

	IR	MU	MD	M_E	M_2	M_D	M_{SVD}
Before-3	0.5342	0.3772	0.0886	0.6987	0.5780	0.9238	0.5333
After-3	0.4405	0.4362	0.1233	0.8393	0.7670	0.9794	0.6191
Before-6	0.3118	0.4988	0.1894	0.8259	0.5051	<1	0.7373
After-6	0.2197	0.5713	0.2091	0.9364	0.7042	1	0.8306

The first observation on Table 8 is that the immobility ratio (IR) decreases after Sandy Hook signifying that users transition between different states more often after than before the event. Specifically, the transitions towards extreme gun rights (MU) increase more than the transitions towards extreme gun control (MD). Thus, overall, the system moves towards extreme stances and mainly exploration of gun rights. The conclusion that the system exhibits more mobility after the event can also be drawn from the eigenvalue-based indices, all of which increase.

The indices described above are used to assess the underlying mobility behaviors in an *individual* transition matrix $\mathbf{P}^n$, but *not* the similarities between different transition matrices. To compute the latter, we need to introduce the notion of comparison between matrices. The first step towards this goal is to have both matrices at the same base, which is achieved by computing their deviation from a perfectly immobile system described by the identity matrix $\mathbf{Q} = \mathbf{I}$. Among the matrix distances in the literature, we use:

- the L1-norm $||\mathbf{P} - \mathbf{Q}||_1 = \sum_i \sum_j (p_{ij} - q_{ij})$,
- the L_2 norm $||\mathbf{P} - \mathbf{Q}||_2 = \sqrt{\sum_i \sum_j (p_{ij} - q_{ij})^2}$,
- the SVD distance $D_{SVD} = M_{SVD}(\mathbf{P}) - M_{SVD}(\mathbf{Q})$, where M_{SVD} is the SVD index defined above, and
- two "risk"-adjusted difference indices, D_1 and D_3, which have the advantage of detecting the direction of the transition, while weighing proportionally "close" and "far" transitions by the factor $(i - j)$:

$$D_1(\mathbf{P}, \mathbf{Q}) = \sum_i \sum_j (i - j)(p_{ij} - q_{ij})$$

$$D_3(\mathbf{P}, \mathbf{Q}) = \sum_i \sum_j (i - j) sign(p_{ij} - q_{ij})(p_{ij} - q_{ij})^2.$$

The distances are given in Table 9. The L_1, L_2 and SVD distances show that after Sandy Hook the users transition between different states more often than before, while the distances of $\mathbf{P}_{after}$ from $\mathbf{I}$ are bigger than the distances of $\mathbf{P}_{before}$ from $\mathbf{I}$. That conclusion is also corroborated by the "risk"-adjusted difference indices, which also bear the information that the majority of transitions are towards gun rights webpages, as indicated by the negative values of the indices.

Table 9: *Common* users: Distances of transition matrices from the immobility matrix $\mathbf{I}$. The top rows correspond to transitions between the 3 high-level states. Similarly, the bottom rows correspond to transitions between the 6 expanded states.

	$\mathbf{L_1}$	$\mathbf{L_2}$	**SVD**	$\mathbf{D_1}$	$\mathbf{D_3}$
Before-3	2.7947	1.1386	0.4294	-1.1838	-0.5739
After-3	3.3571	1.31382	0.5174	-1.3384	-0.7197
Before-6	8.2586	2.0464	0.6197	-5.2088	-1.6976
After-6	9.3639	2.2484	0.6958	-6.021	-2.1765

4.3 Are the Balanced Sites Mediators?

Intuitively, one would expect the balanced pages to act as mediators among websites of opposing viewpoints. We are interested in determining whether they are structured to encourage this type of consumption. Is the structure of the web graph such that the balanced pages can be used as jumping-off points?

To this end, we enumerate the direct transitions among gun control and rights websites. The number of indirect transitions via a balanced webpage is complementary. The percentages of the direct transitions before, and after Sandy Hook are given in Table 10.

Table 10: *All* users: Percent of direct transitions between gun control and gun rights webpages.

	Before	**After**	**Overall**
Control → Rights	91.50%	85.40%	86.30%
Rights → Control	82.20%	86.90%	86.00%

We observe that the transitions between pages that support gun control and pages that support gun rights occur mostly in a direct way, *without* accessing a balanced page. Moreover, before Sandy Hook, the percent of *direct* control to rights transitions (91.5%) is bigger than the percent of rights to control (82.2%) transitions, while the opposite holds true after Sandy Hook. In conclusion, we do not see evidence that the balanced web domains serve as bridges between gun control and rights webpages.

5. CONCLUSIONS

We have examined the temporal browsing behavior of searchers for the controversial and polarizing topic of gun control. We focused on the influence of a single disruptive and shocking news event about the tragic massive shooting at the Sandy Hook Elementary School in December 2012. Our study of a large corpus of web logs from November and December 2012 that is related to gun control, showed that (a) people use the web to largely access agreeable information, and (b) half of the transitions are from gun control to gun rights pages. As for the Sandy Hook shootings, they make the system move into extreme stances, and mainly towards content taking an extreme gun rights stance.

We believe that the methods and results shared in this paper represent an initial step in the realm of analyzing log data to understand how people navigate webpages on controversial topics.

6. REFERENCES

[1] Rawia Awadallah, Maya Ramanath, and Gerhard Weikum. Harmony and dissonance: organizing the people's voices on political controversies. In *WSDM*, pages 523–532. ACM, 2012.

[2] Alexander Kotov, Paul N. Bennett, Ryen W. White, Susan T. Dumais, and Jaime Teevan. Modeling and analysis of cross-session search tasks. In *ACM SIGIR*, pages 5–14, 2011.

[3] Danai Koutra, Paul N. Bennett, and Eric Horvitz. Events and controversies: Influences of a shocking news event on information seeking. *CoRR*, abs/1405.1486, 2014.

[4] Vasileios Koutras and Konstantinos Drakos. A migration approach for USA banks' capitalization: Are the 00s the same with the 90s? *IRFA*, 30(0):131 – 140, 2013.

[5] Eli Pariser. *The Filter Bubble: What the Internet Is Hiding from You*. The Penguin Group, 2011.

[6] Paul Resnick, R. Kelly Garrett, Travis Kriplean, Sean A. Munson, and Natalie Jomini Stroud. Bursting your (filter) bubble: strategies for promoting diverse exposure. In *CSCW*, pages 95–100. ACM, 2013.

[7] PM Sommers and J Conlsk. Eigenvalue immobility measures for markov-chains. *J. Math. Sociol.*, 6(2):253–276, 1979.

[8] Ryen W. White, Susan T. Dumais, and Jaime Teevan. Characterizing the influence of domain expertise on web search behavior. In *WSDM*, pages 132–141, 2009.

[9] E. Yom-Tov, S.T. Dumais, and Q. Gao. Promoting civil discourse through search engine diversity. In *QPOL*, 2013.

[10] B. Yu, S. Kaufmann, and D. Diermeier. Classifying party affiliation from political speech. *JITP*, 5(1):33–48, 2008.