Interactive Filter and Display of Hillary Clinton's Emails: A Cautionary Tale of Metadata

Christopher D. Salahub and R. Wayne Oldford

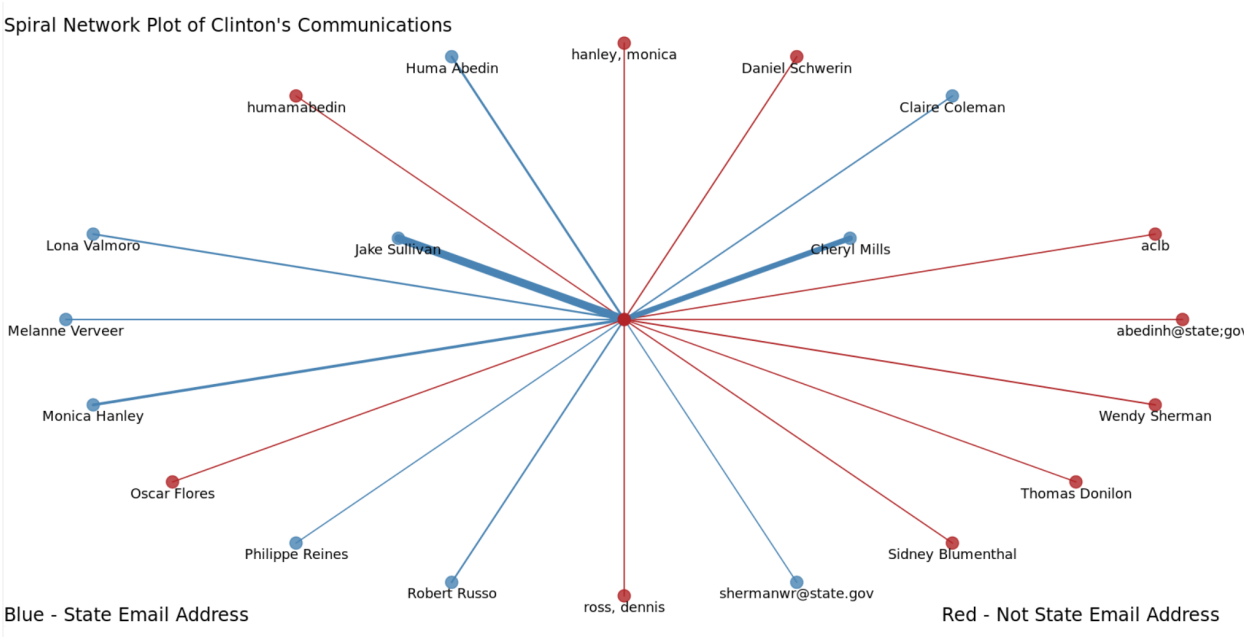

Fig. 1. In the Clouds: Vancouver from Cypress Mountain. Note that the teaser may not be wider than the abstract block.

Abstract—We present a web-based visualization that allows the user to interactively filter and display characteristics of 32,795 Hillary Clinton's emails as provided by Wikileaks.

The visualization focuses on the meta-data of each email, including its senders, receivers, and the timestamp the email appeared on the Clinton server. An interactive time range slider filters all email and all displays automatically update to changes in the slider. The main display shows Clinton's most frequent correspondents arranged as nodes of a spoked graph with Clinton at the centre. Volume determines the thickness of each spoke and high volume determines an inner circle whose spokes are shortened. Correspondents and their edges are coloured according to whether that email account could be identified as being an approved Federal government account or not. A second display shows two daily time series: the total number of emails for that day, and the number meeting selection criteria. A third display shows a scatterplot of the time of day versus the day on which that email appeared. Scatterplot points are coloured by whether the email was redacted or not.

Other displays add some information beyond metadata. FOIA exemption codes appear as a selectable list and a barplot shows email counts by FOIA code. The (stemmed) terms having highest frequency in the displayed email, and those having highest tf-idf are listed in separate displays. All displays are interactively filtered by time range and selected FOIA codes.

We illustrate how the filtered displays can be used to generate hypotheses and uncover interesting information. These touch on contentious issues including the handling of classified information, the 2012 attack on the Benghazi U.S. diplomatic compound, and emails apparently missing from those released publicly.

The data are extracted from Wikileaks HTML files, cleaned, and stored in a form useful for interactive exploration. A local `R shiny` server provides the interactive displays as a public service online tool to explore and uncover patterns in the meta-data and summary contents of Clinton's email. Coupled with publicly available sources of information, these interactive tools uncover surprising amounts of information about an individual, especially one holding public office. The ease with which this can be accomplished and shared should serve as a clear warning as to what can be learned about anyone from metadata.

Index Terms—Exploratory data analysis, metadata, text mining, web-scraping, interactive web visualization, `R`, `shiny`

1 Introduction

The 2016 US Presidential election was one of the most contentious in history. The existence and possible content of Hillary Clinton's private email server played an significant role in the campaign until the end.

Historically important, possibly had an impact on the 2016 US presidential election.

- *Christopher D. Salahub, University of Waterloo, Canada E-mail: csalahub@uwaterloo.ca.*
- *R. Wayne Oldford, University of Waterloo, Canada E-mail: rwoldford@uwaterloo.ca.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org.

Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

Contentious issues:

- private email server
- classified documents (outside of state)
- Sidney Blumenthal
- Benghazi spin handling of media (Susan Rice)
- scrubbing of her server (bleachbit)
- missing email from online email State department

Learn:

- inner circle over time (state or not)
- spike in email around Libyan revolution
- gap in the email
- daily email patterns, server behaviour

Filter:

- time
- redacted or not
- FOIA exemption codes

Content:

- Term frequency, TFIDF

Tool:

- Web-based, interactive filter and display tool

Discovery from visualization & connecting discovery sources
- Nothing on this ... Preparation for Benghazi (security considerations)
- House Oversight and Government Reform Committee (standing committee) Darrell Issa, Chair Jason Chaffetz, Chair (Gowdy a member) (discovered email server) ... - House Select Committee on Benghazi (Summer 2014) Trey Gowdy, Chair

2 Exposition

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat [**?**].

$$\sum_{j=1}^{z} j = \frac{z(z+1)}{2} \tag{1}$$

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Table 1. VIS/VisWeek accepted/presented papers: 1990–2015.

year	Vis/SciVis	SciVis conf	InfoVis	VAST	VAST conf	TVCG @ VIS	CG&A @ VIS	VIS/VisWeek incl. TVCG/CG&A	VIS/VisWeek w/o TVCG/CG&A
2015	33	9	38	33	14	17	15	159	127
2014	34		45	33	21	20		153	133
2013	31		38	32		20		121	101
2012	42		44	30		23		139	116
2011	49		44	26		20		139	119
2010	48		35	26				109	109
2009	54		37	26				117	117
2008	50		28	21				99	99
2007	56		27	24				107	107
2006	63		24	26				113	113
2005	88		31					119	119
2004	70		27					97	97
2003	74		29					103	103
2002	78		23					101	101
2001	74		22					96	96
2000	73		20					93	93
1999	69		19					88	88
1998	72		18					90	90
1997	72		16					88	88
1996	65		12					77	77
1995	56		18					74	74
1994	53							53	53
1993	55							55	55
1992	53							53	53
1991	50							50	50
1990	53							53	53
sum	**1515**	**9**	**595**	**277**	**35**	**100**	**15**	**2546**	**2431**

2.1 Lorem ipsum

Lorem ipsum dolor sit amet (see **??**), consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum.

2.2 Mezcal Head

Lorem ipsum dolor sit amet (see **??**), consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

2.2.1 Duis Autem

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur

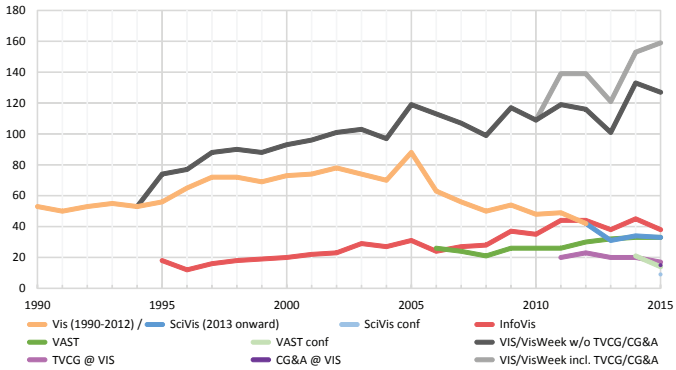

Fig. 2. A visualization of the data from **??**. The image is from [**?**] and is in the public domain.

sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est. Lorem ipsum dolor sit amet.

2.2.2 Ejector Seat Reservation

Duis autem [**?**][1] vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat,[2] vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat.

Confirmed Ejector Seat Reservation Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat [**?**]. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi.

Rejected Ejector Seat Reservation Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie

3 Conclusion

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum.

Acknowledgments

The authors wish to thank A, B, C. This work was supported in part by a grant from XYZ.

[1] The algorithm behind Marching Cubes [**?**] had already been described by Wyvill et al. [**?**] a year earlier.

[2] Footnotes appear at the bottom of the column.