

Investigating and Forecasting User Activities in Newsblogs: A Study of Seasonality, Volatility and Attention Burst

Rafet Sifa

Rafet.Sifa@iais.fraunhofer.de

Fraunhofer Institute for
Intelligent Analysis and Information Systems IAIS

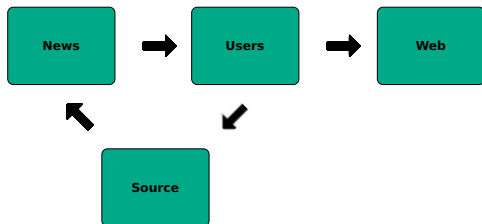
June 12, 2017



- Introduction and motivation to Web-user Analytics
- A look at Wordpress
- Periodicity analysis
- Attention shifts
- Understanding system volatility and predictive analysis

Introduction

- Most of the attention: News Virality and User Behavior
- How about the sources
- Do they have a saying?

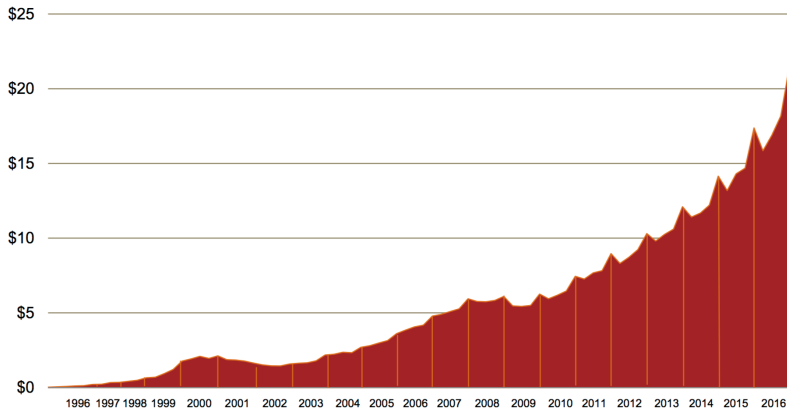


Introduction (2)

- Studying behavioral activities of newsblog users with respect to seasonality, attention shifts and volatility
- Aiming to find possible patterns to help content provider and advertisement operators with informed decision making
- Our proxy: **Number of comments** made on posts

Introduction (3)

Quarterly revenue growth trends 1996-2016 (\$ billions)



Source: IAB/PwC Internet Ad Revenue Report, FY 2016



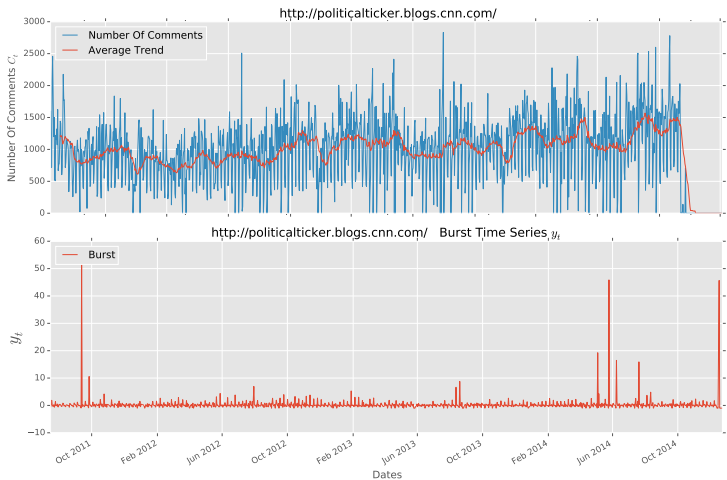
The screenshot shows the WordPress dashboard interface. At the top, the site name 'Restaurant World You...' is visible on the left, and navigation links for 'Upgrade to Pro', 'New Post', 'Dave', and 'Help' are on the right. A dark sidebar on the left contains a menu with items like 'Dashboard', 'Home', 'Comments I've Made', 'Site Stats', 'Akismet Stats', 'My Blogs', 'Blogs I Follow', 'Store', 'Posts', 'Media', 'Links', 'Pages', 'Comments', 'Feedbacks', 'Appearance', 'Users', 'Tools', 'Settings', and 'Collapse menu'. The main content area is titled 'Dashboard' and features several widgets. The 'Right Now' widget displays content and discussion statistics: 8 Posts, 1 Page, 5 Categories, 52 Tags, 9 Comments, 9 Approved, 0 Pending, and 0 Spam. Below this is a message from Akismet stating it has protected the site from 786 spam comments. The 'STORAGE SPACE' widget shows 3,072MB space allowed and 0.08MB (0%) space used. The 'Recent Comments' widget shows two comments from users 'Dave' and 'Mandy' on a post titled 'Arctic Char'. On the right side, the 'QuickPress' widget prompts the user to try a new home page quick post form, with fields for 'Enter title here', 'Add Media', and 'Tags', and buttons for 'Save Draft', 'Reset', and 'Publish'. Below it, the 'Recent Drafts' widget shows 'There are no drafts at the moment', and the 'Stats' widget shows 'No stats are available for this time period.'

Data Set

number of blogs	203
number of posts	713,122
number of comments	14,883,752
time span of activity	2006-2014

Blog URL	News Genre	Viral Post Count
le-grove.co.uk	Sport	620
politicalticker.blogs.cnn.com	Politics	542
order-order.com	Politics	294
sloone.wordpress.com	Personal	248
technologizer.com	Technology	136
snsdkorean.wordpress.com	Entertainment	116
collegecandy.com	Magazine	108
seokyualways.wordpress.com	Personal	96
religion.blogs.cnn.com	Religion	93
kickdefella.wordpress.com	Personal	89

A Typical Newsblog Time Series



Modeling Burst and Volatility

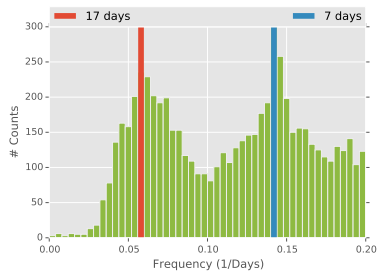
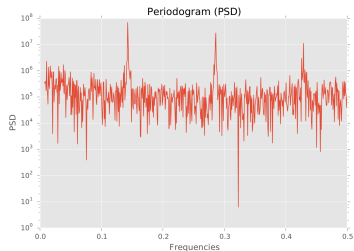
In order to characterize the popularity of a particular news source, we require a quantitative measurement of attention shifts. These can be described via an iter-burst (logarithmic derivative) measure:

$$y_t = \frac{C_t - C_{t-1}}{C_t}$$

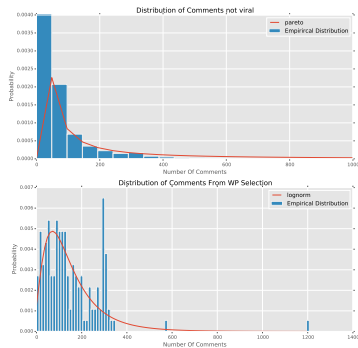
For volatility we refer to Generalized Autoregressive Conditional Heteroscedasticity (GARCH):

$$\sigma_t^2 = \omega + \sum_{k=1}^q \alpha_k \epsilon_{t-k}^2 + \sum_{l=1}^p \beta_l \sigma_{t-l}^2$$

Periodogram

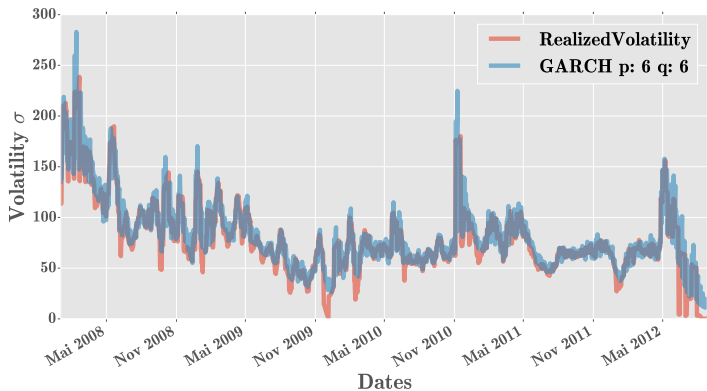


Attention Proxy



- Normal blogs: an average fluctuation value of -96.14
- Relevant blogs: average fluctuation value for the important posts of 71.12
- This shows that *most of the comment behavior arises from popular or viral posts*

Realized and Conditional Volatility



Blogs	Best Model	MSE	MSE $p=1$ $q=1$	# Comments
le-grove.co.uk/	(7, 7)	2.00	1.87	1625633
politickticker.blogs.cnn.com/	(3, 2)	4.48	4.16	1461266
order-order.com/	(5, 5)	4.87	3.80	1289484
sloone.wordpress.com/	(8, 7)	1.39	1.16	93493
technologizer.com/	(9, 1)	594.34	649.79	165856
snsdkorean.wordpress.com/	(2, 5)	0.40	0.28	83081
collegecandy.com/	(9, 1)	0.71	0.61	51495
seokyualways.wordpress.com/	(1, 6)	3.76	2.41	130460
religion.blogs.cnn.com/	(9, 9)	21.66	18.74	2426437
kickdefella.wordpress.com/	(6, 8)	0.22	0.17	39312

- MSE of .74 for best AIC models (100 blogs)
- MSE of .66 for $p=1$ $q=1$ (100 blogs)

Conclusion

- We aimed at understanding the dynamics of the collective attention of populations of newsblogs readers
- Seasonality is uncovered by the analysis of the periodogram function of the time series of comments
- We used logarithmic derivative to model attention shifts and found that virality plays an important role for the shifts
- Volatility, on the other hand, is uncovered by analyzing the deviation of the variable fluctuations in time
- Found behavioral patterns of news-reader attention
- **Future Work:** Incorporating more information about the users and the content into attention modeling

Thank You!

rafet.sifa@iais.fraunhofer.de

<https://mmprec.iais.fraunhofer.de/team/rafet-sifa>



Fraunhofer

IAIS