Dynamics of Media Attention

Traag, V.A.; Reinanda, R.; Hicks, J.; van Klinken, G.

Publication date
2014

Document Version
Final published version

Published in
ICCSA 2014: Fourth International Conference on Complex Systems and Applications: June 23-26, 2014, Le Havre, Normandie, France

ICCSA 2014

Fourth International Conference on Complex Systems and Applications

June 23-26, 2014
Le Havre, Normandie, France

Editors
M.A. Aziz-Alaoui, Cyrille Bertelle,
Xinzhi Liu and Damien Olivier

DYNAMICS OF MEDIA ATTENTION

V.A. Traag*, R. Reinanda, J. Hicks, G. Van Klinken †‡§

Abstract. Studies of human attention dynamics analyses how attention is focused on specific topics, issues or people. In online social media, there are clear signs of exogenous shocks, bursty dynamics, and an exponential or powerlaw lifetime distribution. We here analyse the attention dynamics of traditional media, focussing on co-occurrence of people in newspaper articles. The results are quite different from online social networks and attention. Different regimes seem to be operating at two different time scales. At short time scales we see evidence of bursty dynamics and fast decaying edge lifetimes and attention. This behaviour disappears for longer time scales, and in that regime we find Poissonian dynamics and slower decaying lifetimes. This suggests that a cascading Poisson process may take place, with issues arising at a constant rate over a long time scale, and faster dynamics at a smaller time scale.

Keywords. co-occurrence network • media attention • attention dynamics • lifetime • Poisson process.

1 Introduction

With the arrival of large scale data sets, interest in quantifying human attention rose. It became possible to measure quite precisely how attention grew and decayed [1]. Moreover, it appeared that many human dynamics showed signs of bursty behaviour: short windows of intense activity with long intermittent time spans of inactivity [2]. The duration a person is active—the time between its first and last occurrence, i.e. its lifetime—seems to decay as an exponential, while the edge lifetime seems to follow a powerlaw distribution [3, 4].

We analyse a large dataset of newspaper articles from traditional printed media. We show that the dynamics of this dataset are quite different from social media. Our data consist of 140 263 newspaper articles from Indonesia from roughly 2004 to 2012, gathered by a news service called *Joyo*, mainly focussing on political news. We automatically identify entities by using a technique known as named entity recognition, and only retain person names (we discard organisations and locations) [5]. We then construct a network by creating a node for every person and an edge for each co-occurrence between two persons, and we record the date of the co-occurrence. We only take into account co-occurrences of people in the same sentence, and only about 3.2% of the sentences contain more than one person, so this is quite restrictive. All time is measured in days.

*Corresponding author: `vincent@traag.net`
†All authors are with the KITLV, Leiden, the Netherlands
‡R. Reinanda is associated to Faculty of Sciences, University of Amsterdam, the Netherlands
§Manuscript received April 19, 2009; revised January 11, 2010.

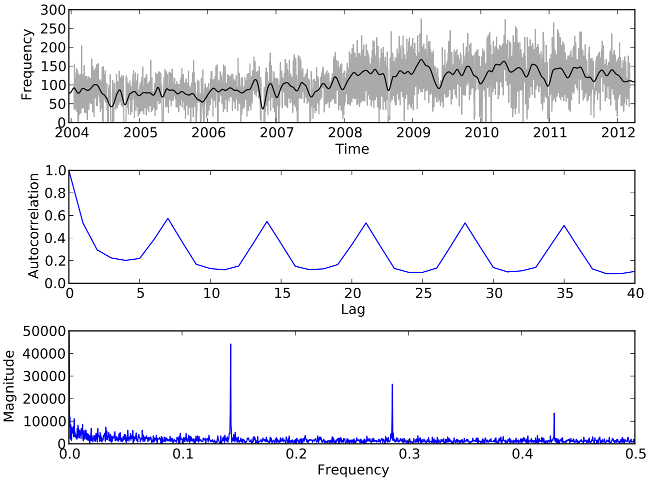


Figure 1: Daily node frequency. The first panel contains the number of nodes in the newspaper (that have at least one co-occurrence) per day (in gray), with the solid black line representing a smoothing (using convolution with a Hann window of 8 weeks) of this daily frequency. The second panel shows the autocorrelation function, which shows a clear peak at a lag of 7 days, indicating there is a weekly pattern in the data. This is also confirmed by Fourier analysis in the third panel, which shows a clear peak at a frequency of $1/7 \approx 0.14$.

2 Results

In total, there are $n = 9\,467$ nodes and they have about $\langle k_i \rangle \approx 12$ neighbours on average. Two people co-occur on average about 3 times. Let us first simply look at how these quantities vary over time. Let E_t be the number of co-occurrences at time t, and N_t the number of nodes that have a co-occurrence at time t. The dynamics of N_t follow a distinctive weekly pattern (Fig. 1). This is confirmed by the autocorrelation function, which shows a clear peak at a lag of 7 days with a correlation of about 0.57, while the Fourier transform shows clear peaks at a frequency of about $1/7 \approx 0.14$. Results for E_t follow a similar pat-

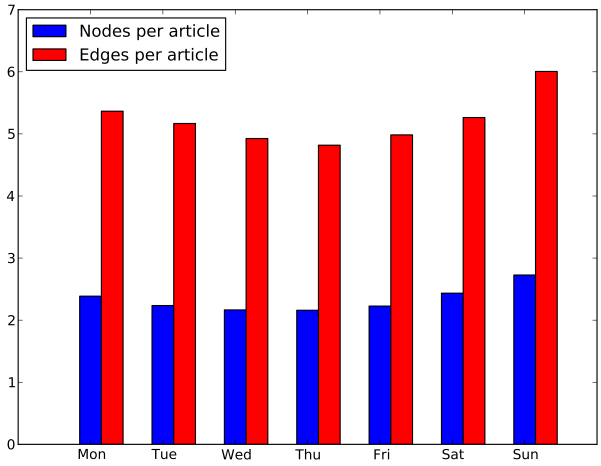


Figure 2: Number of nodes and edges per article per day of the week. Although the largest part of the cyclical behaviour is due to the weekly newscycle (weekend vs. weekday), there still remains some cyclical patterns after normalisation.

tern. Although this largely follows the weekly cycle of the number of articles, some cyclic pattern remains if the node frequencies are normalised by the article frequencies (Fig. 2). For the nodes this pattern is quite weak, but for the edges more noticeable. Surprisingly, the cycle seems to be at its high point at the end/beginning of the week, while its low point is attained in the middle of the week.

Let us denote by $N_t(i)$ the number of co-occurrences of node i with any other node at time t. Neighbours tend to show quite similar patterns of attention, much more than compared to the overall trend. This suggests that attention for people rises and falls together, hinting at some underlying commonality. One possible explanation is that issues arise in which people play a role together, so that they tend to show similar patterns of attention.

Let $t_p(i) = \arg\max_t N_t(i)$ be the peak of the number of co-occurrences of node i with any other node (if there are multiple such times the first is used). We then normalise the time series, such that the peak is centred at 0 with a value of 1, and denote the average of these time series by $\tilde{N}_t$ (see Fig. 3a). Hence $\tilde{N}_0 = 1$, and we are interested in how $\tilde{N}_t$ grows for $t < 0$ and decays for $t > 0$. It was suggested that $\tilde{N}_t \sim |t|^{-\beta}$, where different exponents β would correspond to different universal classes [1]. However, it was also suggested that $\tilde{N}_t \sim -\beta \log |t|$, which is unrealistic for large times since $-\log |t| < 0$ for sufficiently large t. Alternatively, an exponential growth and decay $\tilde{N}_t \sim e^{-\beta|t|}$ is a common phenomena, suggesting the rate of growth/decay is constant throughout time.

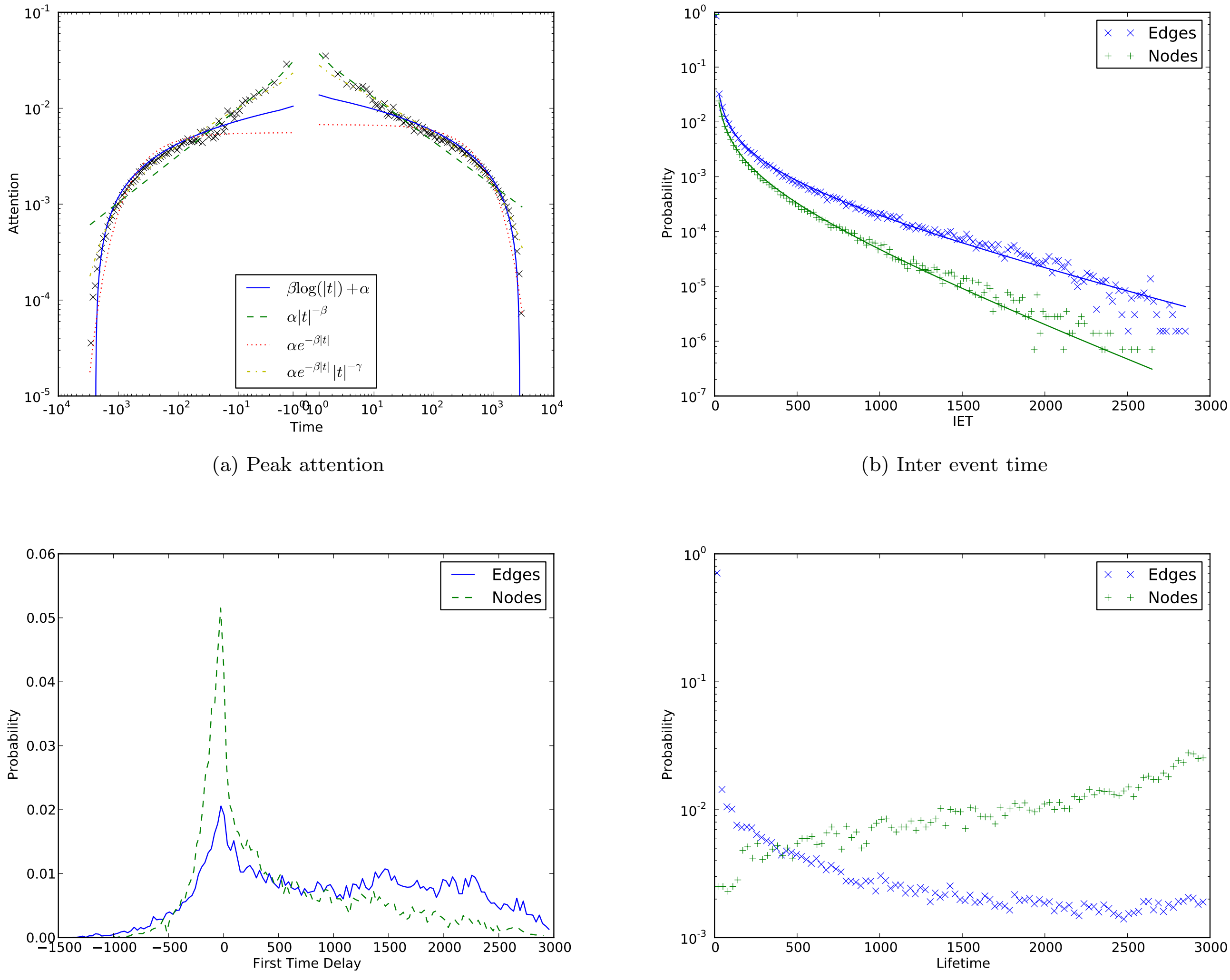
However, we find that none of these satisfactorily model the growth and decay of attention. The logarithmic and exponential grow/decay too slowly at a small time (around 10–30 days), while the powerlaw poorly fits the dynamics for larger times. This suggests that $\tilde{N}_t \sim e^{-\beta|t|}|t|^{-\gamma}$ might be a better fit. Indeed, this functional form quite accurately captures the growth and decay in our data, and comparing AIC values, clearly points to this model (see Table 1 for results). This suggests that at a short timescale there is a very fast (powerlaw) decay, but that at longer timescales the exponential decay dominates. In fact, it suggest that the attention essentially diverges for $|t| \to 0$.

Although there is a large degree of symmetry, which would suggest an endogenous pattern following [1], we do find different exponents for the growth and decay (see Table 1). In particular, the decay seems slightly slower than the growth at short time scales, and nodes tend to occur more frequently after their peak than before their peak on longer time scales, contrary to [6].

Let $t_s(i,j)$ be the time of the s^{th} co-occurrence between node i and j. The inter-event time can then be denoted by $\delta_s(i,j) = t_{s+1}(i,j) - t_s(i,j)$, which can possibly be 0 if two (or more) co-occurrences happened at the same day. Similarly for nodes, we denote by $t_s(i)$ the s^{th} co-occurrence of i with another node, and by $\delta_s(i) = t_{s+1}(i) - t_s(i)$ the inter-event time. If events happen at a constant rate, we would expect an exponential distribution of inter-event times. If events follow a bursty patter however, we expect to find a powerlaw inter-event time distribution, often observed in other settings [2]. We find that although the inter-event times decay quite fast at a relatively short time scale, the inter-event times for a larger time scale follow an exponential distribution (Fig. 3b). Altogether, a powerlaw with exponential cutoff $p(x) \sim x^{-\alpha} e^{-\beta x}$ provides the better fit , compared to a pure powerlaw or pure exponential distribution (log-likelihood ratios $6.3 \cdot 10^4$ and $3 \cdot 10^4$). For the edges, using MLE we find coefficients $\alpha \approx 1.003 \pm 2.6 \cdot 10^{-3}$ and $\beta \approx 0.00151 \pm 1.1 \cdot 10^{-5}$, while for the nodes the distribution is slightly less skewed, but decays slightly faster, with $\alpha = 1.045 \pm 3.7 \cdot 10^{-3}$ and $\beta = 0.00244 \pm 2.3 \cdot 10^{-5}$, both using $x_{\min} = 10$. Indeed, for even shorter time intervals, the decay is very fast, and using lower $x_{\min}$ gives poorer fits.

One possible explanation is that if somebody is involved in an issue, his co-occurrence will show a bursty pattern at this shorter time scale, but that the rate at which somebody is involved in an issue follows a Poisson process over a longer time scale, suggesting something like a cascading Poisson process [2].

If the processes would be stationary, then the expected first time we observe a node would be $\langle t_0(i) \rangle = \frac{\langle \delta_s(i)^2 \rangle}{2\langle \delta_s(i) \rangle}$ (and similarly so for edges). Of course there is some censoring in the data, and for those nodes (and edges) we would expect to find $t_0(i) \approx \langle t_0(i) \rangle$. However, if the actual first time a node appears in the paper $t_0(i)$ is much larger than this expected value $\langle t_0(i) \rangle$, it suggest this really is the first time this person appears in the media. We call the difference $\Delta t_0(i) = t_0(i) - \langle t_0(i) \rangle$ the "first time delay", and the distribution is plotted in Fig. 3c. First,



(a) Peak attention

(b) Inter event time

(c) First time delay

(d) Life time

Figure 3: Attention statistics. Panel (a) shows how attention—as in the frequency of occurring—grows and decays with the peak attention centred at zero. Panel (b) shows the inter-event time distribution, which has a clear exponential tail, suggesting a Poisson process. Panel (c) shows the distribution of the first time delay. This delay represents the difference between the time at which we would expect to first observe a node (or edge) and the time at which we actually observed it in the dataset. This suggests that new nodes and edges are continuously being observed. Finally, in panel (d) the lifetime—the amount of time between the first and last time of observing that node or edge—follows a very peculiar distribution. The edge lifetime decays very fast for the first few days, but then only decays very slowly. The node lifetime actually shows an increasing distribution, suggesting they can have a very long lifetime.

there is a clear peak around $\Delta t_0(i) = 0$, suggesting that censoring played a role in observing these nodes for the first time. The fast decay that is visible on the negative part, does not show at the positive part. This implies that many nodes and edges are first observed much later than expected. This suggests that these nodes and edges are only introduced at a later time. Hence, there are continuously new edges and nodes appearing in the news. This seems especially prominent for the edges, which show an almost uniform distribution between 250–2 000 days of difference, whereas the distribution for the nodes decays more continuously.

We denote by $\Delta(i,j) = \max t_s(i,j) - \min t_s(i,j)$ the lifetime of an edge, and similarly by $\Delta(i)$ the lifetime of a node (Fig. 3d). Most edges have a very low lifetime of only a single day, and the probability to have a larger lifetime quickly decreases. Nonetheless, after an initial rapid decay, the probability decays much slower. This

	Model	$\alpha - \beta \log \lvert t \rvert$	$\alpha \lvert t \rvert^{-\beta}$	$\alpha e^{-\beta t}$	$\alpha e^{-\beta \lvert t \rvert} \lvert t \rvert^{-\gamma}$
Growth	α	$0.011 \pm 6.9 \cdot 10^{-5}$	$0.031 \pm 3.2 \cdot 10^{-4}$	$0.0056 \pm 6.3 \cdot 10^{-5}$	$0.0233 \pm 1.9 \cdot 10^{-4}$
	β	$0.0013 \pm 9.8 \cdot 10^{-6}$	$0.49 \pm 2.3 \cdot 10^{-3}$	$0.0019 \pm 3.1 \cdot 10^{-5}$	$6.6 \cdot 10^{-4} \pm 1.2 \cdot 10^{-5}$
	γ	–	–	–	$0.36 \pm \cdot 2.3 \cdot 10^{-3}$
	ΔAIC	3 117	3 045	4 883	–
Decay	α	$0.014 \pm 8.1 \cdot 10^{-5}$	$0.037 \pm 3.8 \cdot 10^{-4}$	$0.0067 \pm 7.1 \cdot 10^{-5}$	$0.028 \pm 2.3 \cdot 10^{-4}$
	β	$0.0017 \pm 1.1 \cdot 10^{-5}$	$0.46 \pm 2.1 \cdot 10^{-3}$	$0.0015 \pm 2.3 \cdot 10^{-5}$	$5.5 \cdot 10^{-4} \pm 9.8 \cdot 10^{-6}$
	γ	–	–	–	$0.34 \pm 2.2 \cdot 10^{-3}$
	ΔAIC	2 559	2 986	4 820	–

Table 1: Growth and decay estimates and model performance. The AIC differences are quite high, and suggest that $\alpha e^{-\beta|t|}|t|^{-\gamma}$ is the best model.

suggests that besides the more volatile short term links, there are quite some long term stable links. The node lifetimes show a quite unexpected behaviour. Although there are nodes that have a lifetime that is quite short, nodes tend to have a longer lifetime, only cutoff by the duration of the dataset. This suggests that the lifetime of nodes can be extremely long, and can easily run in the decades.

3 Conclusion

Online social media show signs of exogenous shocks, bursty dynamics and exponential lifetime distributions. We have shown here that traditional media seems to operate differently. The current results suggest the media operates at two different time scales. There is a short time scale which operates at the level of issues: links have short lifetime, attention decays quickly and there are indications of burstiness. However, at a longer time scale, these issues seem to occur at a uniform rate, and often involve similar actors: nodes and links have a relatively long lifetime, attention decays slower, and inter-event times decay exponentially. We aim to further analyse this idea in future research, following the cascading Poisson model [2].

References

[1] R. Crane, D. Sornette, *Proc. Natl. Acad. Sci. USA* **105**, 15649 (2008).

[2] R. D. Malmgren, D. B. Stouffer, A. E. Motter, L. A. N. Amaral, *Proc. Natl. Acad. Sci. USA* **105**, 18153 (2008).

[3] C. A. Hidalgo, C. Rodriguez-Sickert, *Physica A* **387**, 3017 (2008).

[4] J. Leskovec, L. Backstrom, R. Kumar, A. Tomkins, *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08 (ACM, New York, NY, USA, 2008), p. 462470.

[5] J. R. Finkel, T. Grenager, C. Manning, *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05 (Association for Computational Linguistics, Stroudsburg, PA, USA, 2005), p. 363370.

[6] J. Leskovec, L. Backstrom, J. Kleinberg, *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09 (ACM, New York, NY, USA, 2009), p. 497506.