Dynamics of Media Attention
Traag, V.A.; Reinanda, R.; Hicks, J.; van Klinken, G.A.

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

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
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, focusing 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 focusing 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.

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.

...t is unrealistic for large times since exponential grow/decay too slowly at a small 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 250–2000 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 | $\alpha - \beta \log |t|$ | $\alpha |t|^{-\beta}$ | $\alpha e^{-\beta t}$ | $\alpha e^{-\beta |t| |t|^{-\gamma}}$ |
|-------|-----------------|-----------------|-----------------|-----------------|
| Growth | $\alpha$ | $0.011 \pm 6.9 \cdot 10^{-5}$ | $0.031 \pm 3.2 \cdot 10^{-5}$ | $0.0056 \pm 6.3 \cdot 10^{-5}$ | $0.0233 \pm 1.9 \cdot 10^{-4}$ |
| | $\beta$ | $0.0013 \pm 9.8 \cdot 10^{-6}$ | $0.49 \pm 2.3 \cdot 10^{-3}$ | $0.0019 \pm 3.1 \cdot 10^{-5}$ | $6.6 \cdot 10^{-4} \pm 1.2 \cdot 10^{-5}$ |
| | $\gamma$ | $- $ | $- $ | $- $ | $0.36 \pm 2.3 \cdot 10^{-3}$ |
| $\Delta \text{AIC}$ | $3117$ | $3045$ | $4883$ | $- $ |
| Decay | $\alpha$ | $0.014 \pm 8.1 \cdot 10^{-5}$ | $0.037 \pm 3.8 \cdot 10^{-4}$ | $0.0067 \pm 7.1 \cdot 10^{-5}$ | $0.028 \pm 2.3 \cdot 10^{-4}$ |
| | $\beta$ | $0.0017 \pm 1.1 \cdot 10^{-5}$ | $0.46 \pm 2.1 \cdot 10^{-3}$ | $0.0015 \pm 2.3 \cdot 10^{-5}$ | $5.5 \cdot 10^{-4} \pm 9.8 \cdot 10^{-6}$ |
| | $\gamma$ | $- $ | $- $ | $- $ | $0.34 \pm 2.2 \cdot 10^{-3}$ |
| $\Delta \text{AIC}$ | $2559$ | $2986$ | $4820$ | $- $ |

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 events: 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


