9^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2023, Copenhagen, Denmark

A dynamic perspective on online news article reading

Keywords: online news, log data, feedback loops, traffic sources

Extended Abstract

In contemporary online news environments, users typically consume news *per item* – a development that has been referred to as “unbundling”, “debundling”, or “atomization” (e.g., [1]) This has sparked the interest of both scholars and practitioners, who want to understand which articles are read by whom. Central foci were selective exposure[7] and the effect of audience metrics on journalistic decisions [3].

We add a dynamic perspective to earlier studies on article popularity. First, popularity can be understood as real-time audience feedback, or, simply put, the number of clicks. For instance, many news sites use “popularity-based filtering” to highlight frequently-clicked [5]. In such situations, popularity at any point in time t_i will positively influence popularity at t_{i+1}. Second, popularity can be understood as an explicit cue given to the readers. For instance, readers tend to regard a “Most Read” label displayed next to an article as a cue for the article’s importance [4]. The dynamic nature of popularity is further affected by boosts such as mobile push alerts, newsletters, and social-media campaigns.

We aim to disentangle the interplay of static and dynamic features in explaining a news item’s popularity: How big is the self-reinforcing effect of popularity, also referred to as a “re-inforcing spiral” [6]? Is this true for all articles, or are there differences? And do articles that gain little traction short after being published stand a chance to make up for this later? We also aim to understand how the role of journalistic decisions – for instance, the decision to publish an article at multiple places, or to specifically highlight it – compares to the role of popularity.

We study such feedback looks using log data from a major newspaper publisher. It contains all clicks on all articles for five regional newspapers spanning $N_{views} = 12,107,413$ views of $N_{articles} = 17,994$, split per minute per traffic source, over the course of 13 weeks.

Selected Results and Conclusions

Measures like placing the article in a specific section, or promoting it via push messages or newsletters indeed increases their views (Figure 2). But interestingly, the effect usually is not long-lasting, contradicting the feedback-loop hypothesis. A notable exception are social media: If an article is spread via social media, it is not only clicked more, but also clicked on for a longer period of time. We hypothesized positive relationships between length and shelf-life (H1, supported), between number of sections and total views (H2, supported), between being on the front page and total views (H3, supported), being highlighted or recommended by the editors (H4, only partly supported). We asked about the shape of viewing trajectories (RQ1, indicating for instance a peak at 2 hours after publication and very little views after 24 hours; and that after on average 38 hours, 90% of all visits have been received, the so-called *shelf-life* [2]). We also found differences between sections (RQ2; pronounced differences in views as well as shelf-life), different channels (RQ3, push messages having a large impact), and influence of early popularity (RQ4, next section).

The dynamic influence of popularity on the viewing trajectory

Our main interest lies in the question *if being popular in the first hours after publication would shape the viewing trajectory*. In this abstract, we focus on the shelf-life. Most articles reach their peak within two hours (see also Figure 2). The total views within these first two hours are positively correlated with shelf-life ($\rho(32499) = .13, p < .001$), but this effect does not hold when controlled for in a multilevel model ($B = .01, p = .608$). However, an article has to *compete* for views with other articles published in the same time frame. To account for this, we calculate a moving-window relative popularity *rank* – which is also a proxy of being highlighted in "most read" boxes. A value of 1 means that the article gathered the most views at that point in time. We then find a different pattern: The *less* popular an article is in the beginning (a higher value indicates a lower ranking), the higher its shelf-life ($\rho(32499) = .38, p < .001$), also in a multilevel model ($B = 2.34, p < .001$) (Table1) .

This conflicts with the initial finding that gathering more views, in the beginning, will expand an article's lifespan. A possible explanation for this discrepancy could be that the relationship between popularity, and shelf life is not linear, but rather quadratic. When we take the mean-centred and squared sum of the first two-hour views it indeed negatively correlates with shelf-life ($\rho(32499) = -0.19, p < .000$). Hence, articles that either gather very little or a lot of views at the start, have a shorter lifespan. Yet, this effect does not hold in the full model($B = -.01, p = .649$). When we apply the same transformation to the average popularity rank within the first two hours, we find a positive correlation instead ($\rho(32499) = .04, p < .001$), which reverses in the full model ($B = -.01, p < .001$).

We conclude that viewing trajectories are indeed partly shaped by a dynamic process in which the popularity of an article has ab influence on the shape of the future click time series – but that this influence does *not* necessarily imply a *positive* feedback loop. Based on additional analyses, we propose that this is mainly explained by a difference between short, breaking news and background stories. In future work, we will incorporate content characteristics of the articles, to better disentangle such differences.

References

[1] Axel Bruns. *Gatewatching and news curation: Journalism, social media, and the public sphere*. New York, NY: Lang, 2018.

[2] Carlos Castillo et al. "Characterizing the Life Cycle of Online News Stories Using Social Media Reactions". In: *17th ACM conference on Computer supported cooperative work & social computing (CSCW '14)*. Baltimore, ML: ACM, 2014, pp. 211–223.

[3] Silke Fürst. "In the Service of Good Journalism and Audience Interests? How Audience Metrics Affect News Quality". In: *Media and Communication* 8.3 (2020), pp. 270–280.

[4] Solomon Messing and Sean J. Westwood. "Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation When Selecting News Online". In: *Communication Research* 41.8 (2014), pp. 1042–1063.

[5] Lynge Asbjørn Møller. "Recommended for You: How Newspapers Normalise Algorithmic News Recommendation to Fit Their Gatekeeping Role". In: *Journalism Studies* (2022), pp. 1–18.

[6] MD Slater. "Reinforcing Spirals: The Mutual Influence of Media Selectivity and Media Effects and Their Impact on Individual Behavior and Social Identity". In: *Communication Theory* 17 (2007), pp. 281–303.

[7] Frederik J Zuiderveen Borgesius et al. "Should we worry about filter bubbles ?" In: *Internet Policy Review* 5.1 (2016).

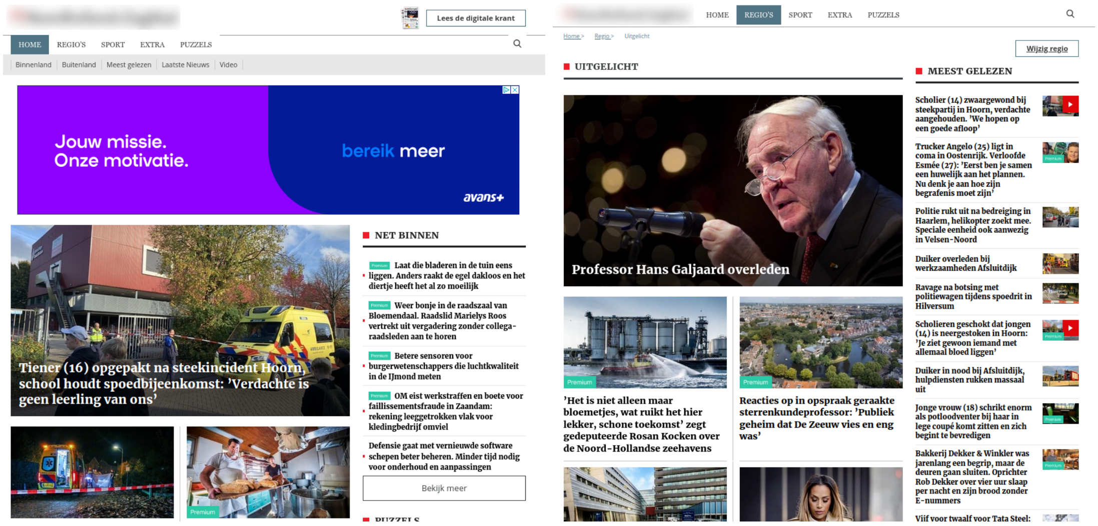

Figure 1: There are multiple ways of accessing the same article: via the homepage (left), or via specific pages and sections like "In the spotlight" ("uitgelicht") or "Most read" ("Meest gelezen", both right). Other pages, which can be selected via the menu bar on top, include specific pages per town, or traditional journalistic sections like "foreign news", "domestic news", "sports", etc. We study how these impact the dynamics of the click time series of an article.

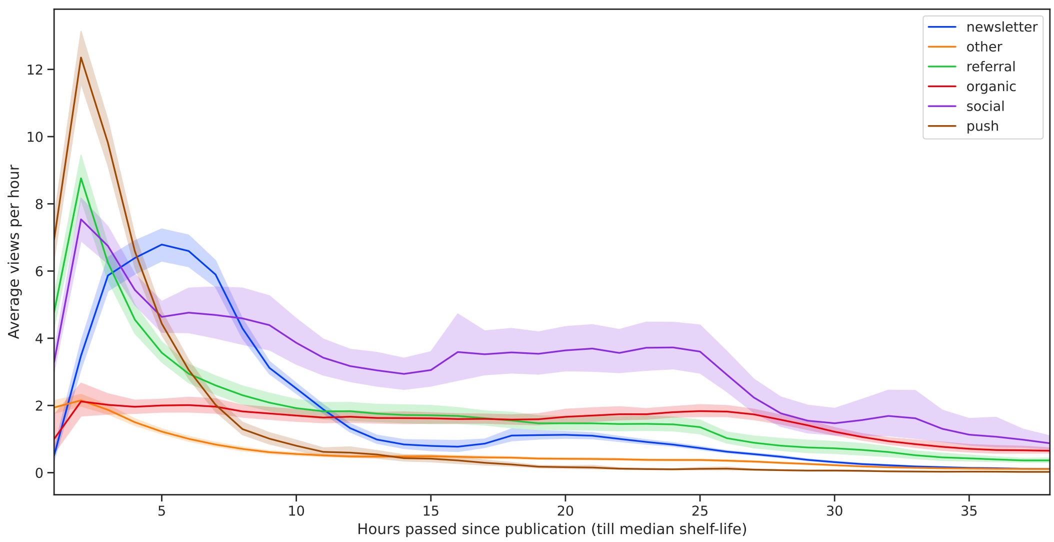

Figure 2: Traffic sources and Viewing trajectory (since publication). Depending on the source, we see differences in short popularity burst versus long-term effects that can extend an article's shelf life.

Table 1: Mixed Linear Model Regression Results

Model:	MixedLM	Dependent Variable:	Shelf_life
No. Observations:	22424	Method:	REML
No. Groups:	7905	Scale:	78434.7220
Min. group size:	2	Log-Likelihood:	-160081.4506
Max. group size:	5	Converged:	Yes
Mean group size:	2.8		

	Coef.	Std.Err.	z	P> |z|	[0.025	0.975]
Intercept	160.222	26.302	6.092	0.000	108.671	211.773
Newspaper[T.Haarlemsdagblad]	-11.083	7.781	-1.424	0.154	-26.333	4.168
Newspaper[T.Ijmuidercourant]	-11.245	6.198	-1.814	0.070	-23.392	0.902
Newspaper[T.Leidschdagblad]	31.004	7.974	3.888	0.000	15.376	46.632
Newspaper[T.Noordhollandsdagblad]	55.005	10.223	5.380	0.000	34.968	75.043
Author[T.editor(s)]	-29.468	20.263	-1.454	0.146	-69.182	10.247
Author[T.individual reporter(s)]	10.975	16.949	0.648	0.517	-22.244	44.194
Author[T.our reporter(s)]	12.467	21.694	0.575	0.566	-30.053	54.987
Author[T.unknown]	-4.570	20.975	-0.218	0.828	-45.681	36.541
Background[T.True]	84.925	19.431	4.371	0.000	46.841	123.010
Culture[T.True]	37.247	25.894	1.438	0.150	-13.504	87.998
Day_of_the_week[T.Monday]	17.978	8.747	2.055	0.040	0.834	35.121
Day_of_the_week[T.Saturday]	-9.441	9.277	-1.018	0.309	-27.623	8.741
Day_of_the_week[T.Sunday]	-22.206	10.051	-2.209	0.027	-41.906	-2.506
Day_of_the_week[T.Thursday]	2.825	8.562	0.330	0.741	-13.956	19.607
Day_of_the_week[T.Tuesday]	4.667	8.565	0.545	0.586	-12.121	21.454
Day_of_the_week[T.Wednesday]	12.076	8.436	1.431	0.152	-4.458	28.610
Domestic[T.True]	-23.063	21.187	-1.089	0.276	-64.590	18.463
Foreign[T.True]	-21.790	21.510	-1.013	0.311	-63.949	20.369
Frontpage[T.True]	-60.708	6.275	-9.675	0.000	-73.007	-48.410
Highlighted[T.True]	-34.810	12.222	-2.848	0.004	-58.765	-10.855
Inline_media[T.True]	-0.400	6.945	-0.058	0.954	-14.011	13.211
Lifestyle[T.True]	66.115	20.013	3.304	0.001	26.890	105.340
Opinion[T.True]	20.897	18.761	1.114	0.265	-15.875	57.668
Paid_content[T.True]	-17.979	8.245	-2.181	0.029	-34.139	-1.819
Recommended[T.True]	80.118	39.611	2.023	0.043	2.481	157.755
Regional[T.True]	63.676	14.102	4.515	0.000	36.036	91.317
Sport[T.True]	19.446	14.665	1.326	0.185	-9.296	48.188
Average_popularity_rank_in_1st_2h	2.344	0.124	18.903	0.000	2.101	2.587
Character_count	0.004	0.001	2.787	0.005	0.001	0.007
Number_of_sections_excl_frontpage	-78.014	18.922	-4.123	0.000	-115.100	-40.927
Time_of_the_day	-2.474	0.514	-4.815	0.000	-3.481	-1.467
Views_in_1st_2h	0.009	0.018	0.513	0.608	-0.026	0.044
Average_popularity_rank_in_1st_2h_sq	-0.013	0.001	-16.491	0.000	-0.014	-0.011
Views_in_1st_2h_sq	-0.000	0.000	-0.456	0.649	-0.000	0.000
Group Var	18348.419	3.280				