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

Diurnal Patterns in the Spread of Misinformation

Keywords: Misinformation, Diurnal Patterns, Cognitive Performance, Social Media, COVID-19

The use of social media as the primary news source and its potential to disrupt democratic institutions and processes has raised concerns regarding spreading false or misleading information. Social media blurs the line between news and personal communication, placing the burden of scrutinising new information for its reliability mainly on the user. Distinguishing between truth and falsehood depends on various factors, including cognitive resources, emotional state, cognitive biases, and prior beliefs. These factors vary between individuals, exhibit regular cyclical behaviours [1], [2], and are impacted by external factors such as light exposure [3] and circadian phase preferences [4].

Mis- and disinformation resist numerous attempts of mitigation. Poor alignment between sleep timing and one's internal circadian rhythm can negatively affect cognitive performance, including attention, memory, decision-making capacity, and reflective thinking [5]. Non-pharmaceutical interventions against the global COVID-19 pandemic, such as lockdowns, curfews, and remote work, may have interacted with circadian mismatches, increasing susceptibility to uncritically believe and disseminate mis- and disinformation. This study investigates how time impacts an individual's propensity to spread mis- and disinformation on Twitter.

In a deductive correlational study, we analysed a secondary Twitter dataset [6] of content containing hashtags and keywords surrounding the COVID-19 pandemic. Tweets containing a link were assigned a source reliability rating based on manually checked web domains. The corresponding content types are outlined in Table 1.

We cluster users based on their average posting activity curve[1] throughout the day. A hierarchical clustering analysis, detailed in Table 2, reveals the presence of three distinct clusters with unique patterns of posting activity. Users with low post rates ($<$ 240 posts) are separated into a fourth cluster. Figure 1 displays each cluster on a polar graph surrounding a 24-hour clock, detailing the posting activity distribution throughout a day. It demonstrates that individuals in the cluster we refer to as *morning types* exhibit their highest activity levels during the early morning hours, while *standard types* demonstrate peak activity levels in the late morning and evening. In contrast, *evening types* display a gradual increase in activity beginning in the afternoon, peaking before midnight. *Infrequent posters* exhibit consistently high levels of activity throughout the day.

Clusters show distinct variation in the content types they spread, as detailed in Table 3[2]. Content type and cluster affiliation are highly dependent ($\chi^2 = 66613$, p-value= .0), with *evening type* users spreading over four times more potentially harmful content than *infrequent posters* and nearly twice as much as *standard type* users. This characteristic is visually evident in Figure 1: Notably, for *evening types*, ratios of potentially harmful content are consistently elevated to the range between .2 and .3 as compared to the other clusters, who overall share a lower proportion of potentially harmful content.

Additionally, we find that the distribution of content type by cluster changes between day and night (Table 5). The proportion of potentially harmful content increases significantly in

[1]The term *posting activity* refers to Tweets, Retweets and Replies.

[2]Each user is assigned a weight that takes into account their total posting activity within the body of analysis due to its statistically significant correlation with the dissemination of potentially harmful content ($\rho = .189$, p-value $= 4.4e^{-20}$, see Table 4 for details). A mathematical definition is given in the notes of Figure 1.

between 6:45 pm and 6:30 am for all clusters (Mann-Whitney $U = 736056$, p-value$= 1.2e^{-04}$). Similarly, more potentially harmful content is spread between sunset and sunrise ($U = 743593$, p-value$= 7.4e^{-04}$). As sleeping schedules can vary independently of sunset or sunrise, we additionally define a period of prolonged wakefulness[3]. Indeed, more potentially harmful content is shared during periods of prolonged wakefulness across all clusters ($U = 585331$, p-value$= 2.2e^{-04}$). Figure 1 visualizes waking time with clock hands in each panel. The coincidence of the period of prolonged waking, shown as the time period enclosed by the clock hands, with the upper 25% quantile of the share of potentially harmful content underlines the connection in between prolonged wakefulness and one's propensity to spread potentially harmful content.

This study's findings highlight the significant impact of a user's total posting activity and posting activity curve throughout the day on their tendency to spread potentially harmful content on Twitter. *Evening types* exhibit a higher propensity for spreading potentially harmful content compared to *morning* and *standard types* as well as *infrequent posters*. Moreover, this tendency was amplified during nighttime and periods of prolonged wakefulness. Our study has important practical implications for users who wish to limit their contribution to the spread of potentially harmful content. Users should be mindful of their Twitter usage during periods of prolonged wakefulness and outside their hours of habitual use to reduce the risk of unintentionally spreading potentially harmful content. On a macroscopic level, these findings indicate both the need for targeted interventions to address the spread of potentially harmful content on social media, and the importance of time and diurnal patterns as a variable.

References

[1] F. Dzogang, S. Lightman, and N. Cristianini, "Circadian mood variations in Twitter content," *Brain and neuroscience advances*, vol. 1, 2017.

[2] R. Gleasure, "Circadian Rhythms and Social Media Information-Sharing," in *Information Systems and Neuroscience*, Springer, 2020, pp. 1–11.

[3] T. Roenneberg, C. J. Kumar, and M. Merrow, "The human circadian clock entrains to sun time," *Current Biology*, vol. 17, no. 2, R44–R45, Jan. 2007.

[4] V. Natale and P. C. Cicogna, "Morningness-eveningness dimension: is it really a continuum?" *Personality and Individual Differences*, vol. 32, no. 5, pp. 809–816, Apr. 2002.

[5] B. I. Oyebode and N. Nicholls, "Does the timing of assessment matter? Circadian mismatch and reflective processing in university students," *International Review of Economics Education*, vol. 38, p. 100 226, 2021.

[6] R. Gallotti, F. Valle, N. Castaldo, P. Sacco, and M. De Domenico, "Assessing the risks of 'infodemics' in response to COVID-19 epidemics," *Nature Human Behaviour*, vol. 4, no. 12, pp. 1285–1293, 2020.

[3] Prolonged wakefulness refers to a period during which a user is awake at a time when they would typically be asleep. Given that prolonged wakefulness can affect a user's cognitive performance [4], we may observe increased levels of potentially harmful content during these periods. A mathematical definition is provided in the notes of Table 5.

Table 1: Classification of social media content adapted from Gallotti, Valle, Castaldo, *et al.* [6]

Category	Potentially Harmful	Characteristics
Science	No	subject to a rigorous validation process by scientific methods
Mainstream media	No	subject to fact checking and media accountability
Satire	No	distorts or misrepresents information for entertainment value, usually is easily identified
Clickbait	No	attempts to pass fabricated to misrepresented information as facts
Other	No	general-purpose category collecting content which is not easily classifiable, includes links that are anonymised and often temporary for higher obscurity (originally “Shadow”), or does not contain links at all.
Political	Yes	aims to build a consensus on a polarised position by omission, manipulation and distortion of information
Fake or hoax	Yes	entirely fabricated or manipulated content that aims to be perceived as realistic and reliable
Conspiracy & junk science	Yes	strongly ideological, inflammatory content alternative or oppositional to tested and accountable knowledge and information with the intent of building echo chambers

Table 2: Statistics for each cluster, resulting from hierarchical clustering. Distances are calculated using Ward’s variance minimization algorithm. The maximum distance within a cluster is indicated in bold font.

	per cluster			distance to		
	posts	users	posts/user	morning[1]	standard[1]	evening[1]
infrequent[2]	7858209	860228	9	2.496	2.481	2.706
morning[1]	3123445	3513	889	**2.496**	3.279	4.257
standard[1]	4219884	4382	963	3.279	**2.362**	3.944
evening[1]	2919309	3349	872	4.257	3.944	**2.706**
bot	63313	25	2533	-	-	-

[1] abbreviate morning, standard, and evening type users, respectively
[2] abbreviates infrequent posters

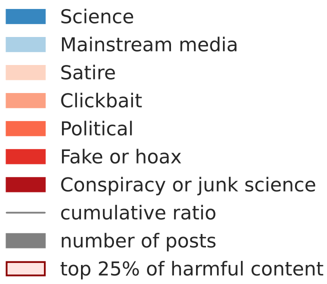
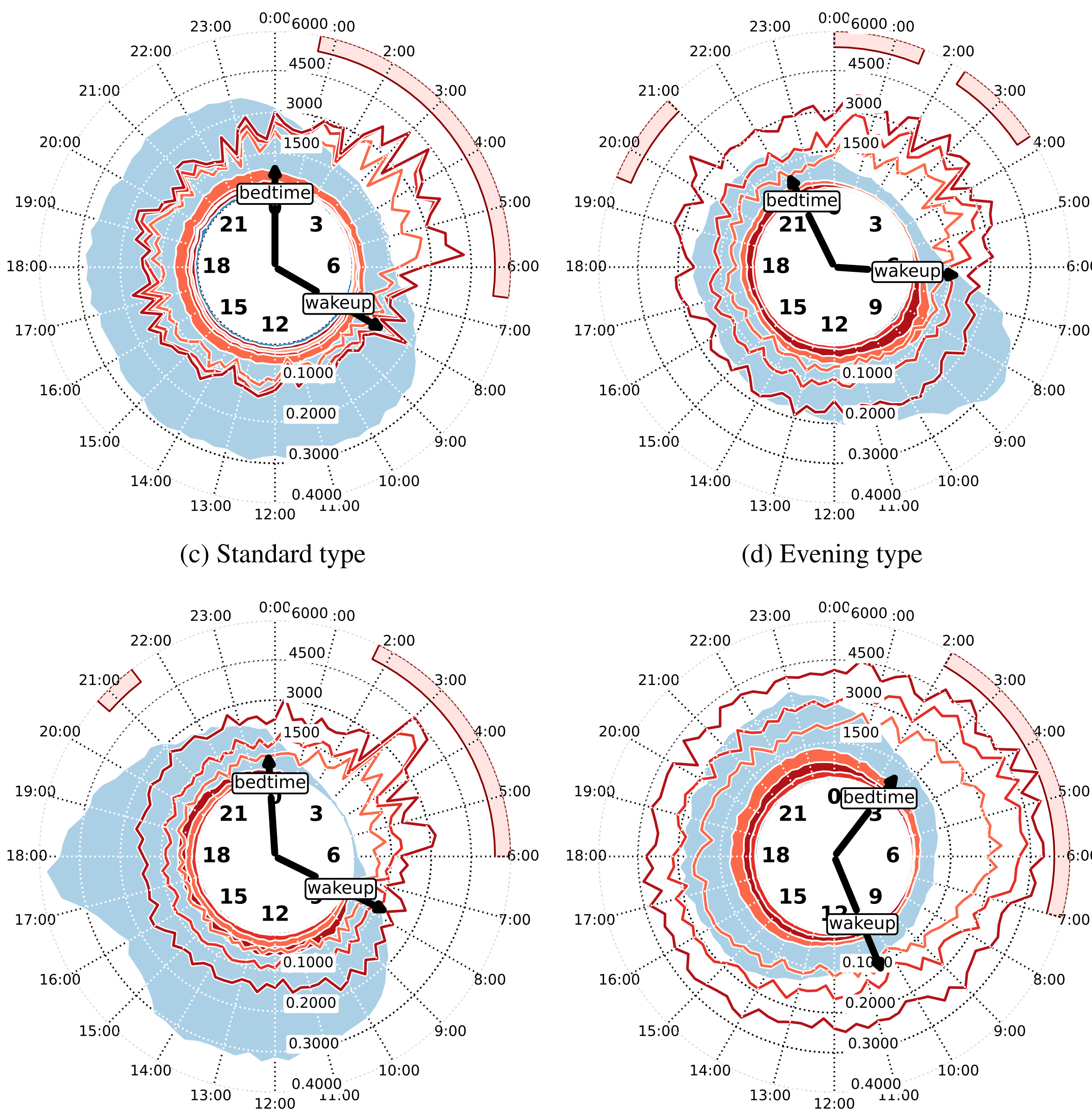


Figure 1: The figure displays, for each cluster: i) the cumulative number of posts with known reliability throughout the day (colored areas), ii) the cumulative ratios[1] of potentially harmful content types, negatively weighted by user activity (lines), iii) the onset ("wakeup") and end ("bedtime") of the user's 16 most active hours[2] (clock hands), iv) the times with the highest quantile of potentially harmful posts (red marginal shaded areas). The axis scales are shared between panels.



(a) Infrequent posters

(b) Morning type

(c) Standard type

(d) Evening type

(Continued on next page)

Figure 1: Continued

[1] Let $T = \{[t, t+\frac{1}{4}) \mid 4t \in \mathbb{N} \wedge 0 \leq t < 24\}$ be the set of 15 minute intervals within a day given in hours, F the set of content types, and I the set of users authoring content. Let then $\{P_{(t,i,f)}\}_{(t,i,f)\in T\times I\times F}$ be the set of posts authored during interval $t \in T$ by user $i \in I$ of content type $f \in F$, indexed by a surjective function from $T \times I \times F$ onto P. Posts are weighted inversely to the total posts per authoring user, with the weight of a given post by user $i \in I$ defined as $w(i) = (\sum_{f\in F}\sum_{t\in T}\left|P_{(t,i,f)}\right|)^{-1}$. Let C be the set of all clusters where $c \in C$ is a subset of I, and F^K the subset of F which are classifiable in terms of reliability (all categories except "Other"). The ratio for content type $f \in F^K$, cluster $c \in C$ and 15 minute time interval within a day $t \in T$ is calculated as $r(t,c,f) = \frac{\sum_{i\in c}\left|P_{(t,i,f)}\right| w(i)}{\sum_{g\in F^K}\sum_{i\in c}\left|P_{(t,i,g)}\right| w(i)}$.

[2] Function $a(t,c) = \frac{\sum_{f\in F}\sum_{i\in c}\left|P_{(i,t,f)}\right|}{\sum_{t\in T}\sum_{f\in F}\sum_{i\in c}\left|P_{(i,t,f)}\right|}$ calculates the activity levels during an interval $t \in T$ by cluster $c \in C$. To denoise the signal, we used the discrete Fourier transform to set all harmonics except those with the three highest amplitudes to 0, leaving us with the smoothed set $\{A_{(t,c)}\}_{(t,c)\in TxC}$. Let $f(t,n) = \{s(\bmod |T|) \mid s \in T \wedge t \leq s < t+n\}$ return the set of 15-minute intervals within $n =$ 16 consecutive hours within a day starting at $t \in T$, where mod refers to the modulo operator. Then, we define the onset of heightened activity as $g(c,n) = \underset{t\in T}{argmax} \sum_{s\in f(t,n)} A_{(s,c)}$, and the end as $(g(c,n)+n)(\bmod |T|)$ for $c \in C$, where $\underset{t\in T}{argmax}\, h(t) = \{t \mid h(x) \leq h(t)\ \forall x \in T\}$ returns the set of points t for which $h(t)$ returns the function's largest value if it exists.

Table 3: Ratios[1] of posts by content type and cluster.

	infrequent posters	morning type	standard type	evening type
Science	0.028	0.021	0.017	0.020
Mainstream media	0.743	0.780	0.820	0.695
Satire	0.008	0.004	0.005	0.004
Clickbait	0.076	0.008	0.005	0.008
Political	0.107	0.078	0.057	0.146
Fake or hoax	0.027	0.040	0.037	0.059
Conspiracy & junk science	0.012	0.069	0.059	0.067
Potentially harmful	0.146	0.187	0.153	0.272

[1] defined in the notes of Figure 1

Table 4: Results of analysis of potentially harmful content spread in relation to cluster activity. The left panel shows the correlation between the ratio of potentially harmful content spread by a user[1] with their total number of posts per user[2]. The right panel displays the correlation between the fraction of activity during a 15-minute time interval within a day[3] and the proportion of potentially harmful content posted during the same interval[4]. The statistics shown in the table are the Pearson correlation coefficient and the corresponding p-value.

	total		per day	
	Pearson R	P-value	Pearson R	P-value
infrequent posters	0.239	1.9e-04	-0.718	1.8e-16
morning type	0.124	2.7e-05	-0.440	7.3e-06
standard type	0.173	2.3e-09	-0.416	2.5e-05
evening type	0.111	1.3e-04	-0.203	4.8e-02
total	0.189	4.4e-20	-0.732	2.3e-17

[1] $\sum_{f \in F^H} \sum_{t \in T} r(t,c,f)$, for a cluster $c \in C$, where F^H consists of the categories of political news, fake or hoax news and conspiracy and junk science. $r(t,c,f)$ is defined in the notes of Figure 1.

[2] $\sum_{f \in F} \sum_{t \in T} \left| P_{(t,i,f)} \right|$ for a user $i \in I$, following the notation defined in the notes of Figure 1

[3] defined in the notes of Figure 1

[4] $\sum_{f \in F^H} r(t,c,f)$, for a cluster $c \in C$ and an interval $t \in T$

Table 5: Mann-Whitney U test comparing the distributions of content type ratios[1] during different time periods: daytime and nighttime, times between and outside of sunrise and sunset, as well as regular and prolonged waking times. P-values that were found to be statistically significant at the $\alpha = .05$ level are highlighted in bold.

		6:30am - 6:45pm[2,3]		sunrise - sunset[2,4]		waking - bedtime[2,5]		lower[6]
		Statistic	P-Value	Statistic	P-Value	Statistic	P-Value	
Potentially harmful	infrequent	718491	**1.4e-05**	729598	**1.9e-04**	567252	**3.3e-05**	day
	morning	722238	**1.8e-04**	710910	**1.9e-05**	628639	**3.6e-02**	day
	standard	611709	**1.2e-06**	626468	**5.5e-05**	422104	**1.3e-04**	day
	evening	700194	**2.5e-07**	684516	**1.4e-09**	625990	**2.4e-02**	day
	total	736056	**1.2e-04**	743593	**7.4e-04**	585331	**2.2e-04**	day
Political	infrequent	635684	**2.7e-09**	651151	**2.1e-07**	520623	**1.3e-04**	day
	morning	536870	**3.8e-13**	554404	**1.0e-10**	421362	**4.0e-15**	day
	standard	405763	**1.6e-15**	423473	**4.3e-11**	171105	**5.5e-26**	day
	evening	573280	**2.7e-17**	583440	**4.5e-15**	498714	**1.5e-08**	day
	total	703391	**2.6e-07**	711236	**2.0e-06**	596405	**1.3e-02**	day
Fake or hoax	infrequent	624548	3.7e-01	629927	2.6e-01	422464	**6.2e-03**	day
	morning	514688	**4.5e-04**	525141	**2.9e-03**	364216	**7.1e-07**	day
	standard	390417	**2.6e-06**	398857	**5.2e-05**	117422	**1.8e-25**	day
	evening	594988	**9.9e-05**	599440	**1.4e-04**	518992	1.0e+00	day
	total	841496	**4.5e-03**	828544	**2.5e-02**	651394	8.5e-02	night
Conspiracy & junk science	infrequent	590097	6.5e-01	575130	4.4e-01	414018	**3.6e-04**	night
	morning	668518	**4.2e-02**	648148	5.7e-01	568820	**1.1e-08**	night
	standard	481363	**1.5e-03**	470474	**1.5e-04**	204671	**3.3e-06**	day
	evening	734568	**2.8e-02**	697408	7.7e-01	582369	**8.4e-04**	night
	total	907976	**1.3e-09**	874539	**2.1e-05**	844999	**4.4e-41**	night

[1] defined in the notes of Figure 1

[2] We account for a safety margin of $s = 1$ hour before and after each border value.

[3] compares the distribution of ratios $r(t,c,f)$ for $t \in$ [7:30 am – 5:45 pm) ("day") with those for $t \in$ [7:45 pm – 6:30 am) ("night"), considering the safety margin. The margin values are sunrise and sunset times averages over the months, rounded to the closest quarter hour.

[4] compares the distribution of ratios sunrise and sunset ("day") with those between sunset and sunrise ("night"). The sunset and sunrise times are calculated geometrically using the average latitude and longitude for the users in our dataset for the first day of each month using Python's suntime library `https://github.com/SatAgro/suntime`.

[5] compares the distributions of ratios within $[(g(c,n)+s)(\bmod |T|), (g(c,n)+n-s)(\bmod |T|))$ ("day") with those of the interval $[(g(c,n)+n+s)(\bmod |T|), (g(c,n)-s)(\bmod |T|))$ ("night") for $n = 16$, where the onset $g(c,n)$ and end of heightened activity $(g(c,n)+n)(\bmod |T|)$ as defined in Figure 1.

[6] For each row, returns the distribution for which the corresponding p value of a one-tailed Mann-Whitney U test was lower for all each significant ($p < \alpha$) comparisons.