

Characterizing Information Diffusion Across Contexts Through Frequency, Diversity, Interaction, and Temporal Dynamics

Jaqueline Faria de Oliveira

jaquefari@gmail.com

Pontifical Catholic University of Minas Gerais

Department of Computer Science

Belo Horizonte, Minas Gerais, Brazil

Instituto Kunumi

Belo Horizonte, Minas Gerais, Brazil

Humberto T. Marques-Neto

humberto@pucminas.br

Pontifical Catholic University of Minas Gerais

Department of Computer Science

Belo Horizonte, Minas Gerais, Brazil

ABSTRACT

Information adoption in online social networks may be preceded by different patterns of social exposure, yet how these patterns vary across information environments remains insufficiently understood. In this paper, we extend a previously proposed ego-network analytical basis into a unified framework for comparative information diffusion analysis. The framework integrates Frequency, Diversity, Interaction, and Temporal Dynamics. The Interaction Dimension is operationalized through the proposed Reinforced Influence Index (RII), which characterizes exposure organization by relating the observed exposure volume to the number of distinct adopting followees. We evaluate the framework using two large-scale ego-network datasets collected from Twitter (now X) during the 2019 European Parliament election in France and the COVID-19 pandemic in Brazil. The analysis is conducted at the item, user, and topic levels to characterize exposure patterns preceding adoption across both datasets. Our results show that repeated and diversified social exposure exhibit distinct patterns across content types, topics, and datasets. RII further reveals differences in exposure organization, including higher User-level exposure-per-source ratios in the Brazilian dataset, while temporal analyses reveal differences in the length and organization of exposure sequences preceding adoption. Overall, the findings reveal both recurring exposure patterns and substantial heterogeneity across analytical dimensions and empirical contexts. The proposed framework provides a unified and interpretable perspective for studying information diffusion and supports comparative analyses across online social systems.

KEYWORDS

information diffusion, threshold models, repeated influence, user heterogeneity, online social networks, ego networks.

1 INTRODUCTION

Information diffusion in online social networks has been extensively studied through global network statistics, probabilistic

diffusion models, and aggregate adoption patterns. Although these approaches successfully characterize macroscopic diffusion, they provide limited insight into the local reinforcement mechanisms that precede adoption from the perspective of individual users. In particular, adoption may be preceded by repeated interactions with the same source or exposure to information from multiple distinct social contacts. Moreover, most empirical studies focus on a single dataset or context, limiting our understanding of how these reinforcement mechanisms vary across different information environments.

Using an ego-network representation of information adoption, we organize information diffusion into four complementary analytical dimensions. Frequency and Diversity characterize repeated and diversified social exposure through Repeated Exposure Influence (REI) and Social Exposure Influence (SEI), respectively, at the *Item*-, *User*-, and *Topic*-levels. We introduce the Reinforced Influence Index (RII), a novel diagnostic indicator that operationalizes the Interaction Dimension by relating the number of observed exposure events to the number of distinct adopting followees, also across these three aggregation levels. Finally, the Temporal Dimension examines how new and repeated exposures are organized over time before adoption.

We evaluate the proposed framework through the first comparative analysis between a previously studied [9, 10] French Twitter dataset collected during the 2019 European Parliament election in France and a newly analyzed Brazilian Twitter dataset on the COVID-19 pandemic in Brazil. These contrasting empirical contexts enable the investigation of diffusion patterns across distinct datasets while preserving a common analytical framework.

This work makes three main contributions. First, we extend our previous ego-network analytical basis into a unified four-dimensional framework that integrates Frequency, Diversity, Interaction, and Temporal Dynamics for comparative analysis. Second, we introduce RII, a novel diagnostic indicator that characterizes exposure organization by measuring the average number of observed exposure events per distinct adopting followee. Third, we conduct the first unified comparison of the previously studied French ego-network dataset and the newly analyzed Brazilian COVID-19 dataset under the same analytical framework.

The comparative analyses show that repeated and diversified social exposure exhibit distinct patterns across topics

In: Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia'2026). Lavras/MG, Brazil. Porto Alegre: Brazilian Computer Society, 2026.

© 2026 SBC – Brazilian Computing Society.
ISSN 2966-2753

and datasets, revealing heterogeneity in the reinforcement patterns preceding adoption. RII further reveals differences in how exposure volume is organized relative to the number of distinct social sources, while temporal analyses reveal differences in how new and repeated exposures are organized over time across the two datasets.

Overall, this work advances the understanding of information diffusion by moving beyond isolated adoption mechanisms and single-context analyses. The proposed framework provides a unified view of Frequency, Diversity, Interaction, and Temporal Dynamics preceding information adoption in online social networks.

2 RELATED WORK

Research on information diffusion has evolved from classical contagion theories to models that examine how social reinforcement, contextual factors, and local network structures shape information adoption. Granovetter [12] first proposed that individuals adopt a behavior once social influence exceeds a personal threshold, an idea later extended by Watts [27] to complex networks, demonstrating how network topology influences cascade dynamics. Building upon these foundations, Centola and colleagues [5–7] distinguished between simple and complex contagions, showing that many behaviors require reinforcement from multiple social contacts rather than a single exposure. Together, these studies established the theoretical foundation for understanding information adoption as a process driven by repeated and diversified social influence.

Subsequent empirical studies demonstrated that different forms of social reinforcement contribute to adoption in complementary ways. Repeated exposure increases adoption through cumulative interactions and memory effects [14, 26, 30], whereas exposure from multiple distinct neighbors provides social validation and reduces uncertainty [22, 29]. Additional work showed that heterogeneous thresholds, temporal dynamics, and network topology further modulate these reinforcement processes [21, 24, 25]. Nevertheless, repeated and diversified exposure are still commonly analyzed independently or treated as a single reinforcement process.

Beyond reinforcement itself, diffusion dynamics have also been shown to vary across information environments and empirical contexts. Studies on political communication, misinformation, crisis events, and public health emergencies have reported differences in adoption behavior associated with topic, urgency, polarization, and temporal structure [2, 4, 16, 18, 19, 28]. Recent work further demonstrates that user heterogeneity, topical domains, and nonlinear reinforcement mechanisms affect cascade dynamics and diffusion outcomes [11, 15, 17].

Ego-network analysis complements these global perspectives by reconstructing adoption from the viewpoint of individual users, enabling the explicit characterization of repeated and diversified exposure before adoption occurs. Classical theories of structural holes and social capital [3], together

with more recent studies on tie strength and structural diversity [1, 13, 15], show that local network organization strongly influences diffusion and user susceptibility.

Despite these advances, important gaps remain. Existing studies rarely integrate repeated exposure, social diversity, their joint organization, and temporal evolution within a unified analytical framework, particularly in comparative analyses across empirical contexts. These gaps motivate the framework proposed in this paper.

3 EXPERIMENTAL SETTING

We evaluate the framework on two large-scale Twitter (now X) datasets: a previously studied dataset from the 2019 European Parliament election in France [9, 10] and a newly analyzed dataset from the COVID-19 pandemic in Brazil. Table 1 summarizes their main characteristics.

Both datasets were reconstructed using the same ego-network methodology [9, 10], including seed-user selection, ego-network reconstruction, and historical timeline collection. Hashtags were grouped into topical clusters using Word2Vec embeddings, t-SNE, and k-means, providing a common basis for Topic-level analyses.

Table 1: Summary of the experimental datasets.

Dataset	Window	Seeds	Users	Tweets	Timelines	Topics
France	May'18–May'19	8,554	1.8M	1.1B	1.2M	10
Brazil	Dec'19–Aug'20	4,237	1.9M	1.5B	1.0M	7

Source: Research data.

4 METHODOLOGY

This work builds upon the ego-network framework and exposure measures introduced in our previous studies [9, 10]. Specifically, the previous framework defined measures of repeated and social exposure, previously termed *susceptibility* and *threshold*, at the Item-, User-, and Topic-levels. In this work, these existing measures are adopted and referred to as Repeated Exposure Influence (REI) and Social Exposure Influence (SEI), respectively, to emphasize their analytical roles as the Frequency and Diversity dimensions.

We extend this previous analytical basis by introducing the Interaction Dimension, operationalized through the Reinforced Influence Index (RII), which relates exposure volume to the number of distinct adopting followees, and by integrating Frequency, Diversity, Interaction, and Temporal Dynamics into a unified framework for comparing information diffusion patterns across the French and Brazilian ego-network datasets.

4.1 Adoption Framework

We adopt the ego-network representation and adoption-event reconstruction defined in our previous studies [9, 10]. In this framework, users are represented by ego-networks composed of their followees. An information item is considered *introduced* when it has previously been published by at least one

followee, and adoption occurs when a user first publishes the same item within a temporal influence window τ after its introduction. The estimation of τ follows the procedure defined in the previous framework. These reconstructed adoption events provide the common methodological basis for the analyses conducted in this work.

The Item-, User-, and Topic-level aggregation scheme is also inherited from the previous framework [9, 10]. The *Item-level* represents individual user-item adoption events, the *User-level* aggregates adoption behavior across all items adopted by a user, and the *Topic-level* specializes the analysis within thematic groups of semantically related items. In the present work, these existing aggregation levels provide a common basis for applying and comparing REI, SEI, and RII, while the Temporal Dimension separately characterizes the organization of exposure sequences preceding adoption.

4.2 Frequency Dimension: Repeated Exposure Influence (REI)

Repeated Exposure Influence (REI), corresponding to the repeated-exposure measure introduced in our previous work [9, 10], characterizes the frequency of repeated exposure preceding adoption.

Item-level. REI is computed for an individual adoption event as

$$S(u, x) = \frac{1}{|X_I(u, x)|}, \quad x \in X_A(u), |X_I(u, x)| \geq 1. \quad (1)$$

User-level. User-level REI aggregates repeated exposure over all adopted items:

$$S(u) = \frac{|X_A(u)|}{|X_I(u)|}. \quad (2)$$

Topic-level. Topic-level REI specializes the same measure within topic X_i :

$$S(u, X_i) = \frac{|X_A(u, X_i)|}{|X_I(u, X_i)|}. \quad (3)$$

Higher REI values indicate that adoption occurs after fewer repeated introductions, whereas lower values correspond to adoption events preceded by a larger number of repeated introductions.

4.3 Diversity Dimension: Social Exposure Influence (SEI)

Social Exposure Influence (SEI), corresponding to the social-exposure measure introduced in our previous work [9, 10], characterizes the diversity of social exposure preceding adoption. Let Fu denote the set of followees of user u , and $F_A u, x$ the subset of followees who adopted item x before u .

Item-level.

$$\phi(u, x) = \frac{|F_A(u, x)|}{|F(u)|}. \quad (4)$$

User-level.

$$\phi(u) = \text{avg}_{x \in X_I(u)} \phi(u, x). \quad (5)$$

Topic-level.

$$\phi(u, X_i) = \text{avg}_{x \in X_I(u, X_i)} \phi(u, x). \quad (6)$$

Higher SEI values indicate that adoption is preceded by exposure from a larger fraction of distinct followees, whereas lower values indicate exposure from a smaller fraction of the user's social neighborhood.

4.4 Interaction Dimension: Reinforced Influence Index (RII)

While REI and SEI characterize complementary properties of the exposure process, neither directly describes how the observed exposure events are distributed across the distinct social sources preceding adoption. The *Interaction Dimension* examines the relationship between these underlying exposure properties by characterizing exposure volume relative to social-source diversity. Rather than representing a statistical interaction between REI and SEI, this dimension captures how observed exposure events are organized across the distinct social sources preceding adoption. To operationalize this relationship, we propose the *Reinforced Influence Index (RII)*, a diagnostic indicator that measures the average number of observed exposure events per distinct adopting followee. RII is derived from the previously defined measures of incoming influence (I_{in}) and unique adopting followees ($k_u \phi$) [9, 10].

The incoming influence and the number of distinct adopting followees are defined in Equations 7 and 8, respectively.

$$I_{in}(u, x) = S(u, x)^{-1}, \quad (7)$$

$$k_u \phi(u, x) = |F_A(u, x)|, \quad |F_A(u, x)| \geq 1. \quad (8)$$

RII is defined at the three aggregation levels as follows.

Item-level.

$$\text{RII}(u, x) = \frac{I_{in}(u, x)}{k_u \phi(u, x)}. \quad (9)$$

Since $I_{in}(u, x) = |X_I(u, x)|$ and $k_u \phi(u, x) = |F_A(u, x)|$, Item-level RII can be equivalently expressed as

$$\text{RII}(u, x) = \frac{|X_I(u, x)|}{|F_A(u, x)|}. \quad (10)$$

This formulation makes the interpretation of RII explicit: it represents the average number of observed exposure events associated with each distinct adopting followee before adoption.

User-level.

$$\text{RII}(u) = \text{avg}_{x \in X_A(u)} \text{RII}(u, x). \quad (11)$$

Topic-level.

$$\text{RII}(u, X_i) = \text{avg}_{x \in X_A(u, X_i)} \text{RII}(u, x), \quad (12)$$

where $|X_A(u, X_i)| \geq 1$ and $|F_A(u, X_i)| \geq 1$.

At the User- and Topic-levels, RII represents the average Item-level exposure-per-source ratio across the corresponding adopted items; it is therefore an aggregation of event-level RII values rather than a ratio computed from pooled exposure and source counts.

RII captures a relational property that is not represented by either REI or SEI alone. REI characterizes exposure frequency without accounting for how those exposures are distributed across social sources, whereas SEI characterizes the fraction of the user's social neighborhood represented by distinct adopting followees without accounting for how many exposure events these sources generate. RII relates these underlying quantities by expressing exposure volume relative to the number of distinct observed sources.

The minimum RII value is 1, which occurs when every observed exposure originates from a different adopting followee. Values greater than 1 indicate that at least some social sources contribute multiple exposure events, with increasing values reflecting a greater concentration of exposures among recurring sources. RII should therefore be interpreted as a descriptive diagnostic of exposure organization preceding adoption, rather than as a causal measure of social influence or an adoption threshold.

4.5 Temporal Dimension: Information Diffusion Dynamics

The temporal descriptors adopted in this dimension build upon the influence-sequence analysis introduced in our previous work [9, 10]. In the present work, they are integrated with REI, SEI, and RII to provide the Temporal Dimension of the comparative framework. While REI, SEI, and RII characterize complementary dimensions of information diffusion, a temporal perspective provides additional insight into how reinforcement unfolds during the adoption process. To this end, each adoption event is represented as an ordered sequence of exposure events, from which three complementary descriptors are derived: influence sequences, sequence entropy, and trigger probabilities.

Influence sequences. For each adopted item, the exposures received by user u before adoption are represented as an ordered sequence $E(u, x) \in \{N, R\}^*$, where N denotes the first exposure originating from a previously unseen followee, whereas R denotes a repeated exposure from a followee already observed in the sequence. By definition, the first event is always labeled N . This representation preserves the temporal ordering of exposure events, providing the basis for analyzing reinforcement dynamics before adoption.

Entropy. To quantify the structural diversity of influence sequences, we employ Shannon entropy. This measure captures how balanced the sequence is between new and repeated exposures. Low entropy indicates reinforcement dominated

by one event type, whereas higher entropy reflects a more heterogeneous combination of both.

$$H(u, x) = \text{Entropy}(E(u, x)). \quad (13)$$

To account for differences in sequence length, entropy is averaged over sequences containing the same number of exposure events:

$$R(n) = \text{avg}(H(u, x))_{|E(u, x)|=n} \quad (14)$$

This normalization enables meaningful structural comparisons independently of the number of observed exposures.

Trigger probabilities. Finally, we estimate the probability that the last exposure preceding adoption originates from a new followee; the complementary probability corresponds to repeated reinforcement.

$$P_N(n) = \Pr(\text{last}(E(u, x)) = N \mid |E(u, x)| = n), \quad (15)$$

$$P_R(n) = 1 - P_N(n). \quad (16)$$

Together, these descriptors characterize the temporal organization of reinforcement preceding information adoption.

4.6 Statistical Comparison

Cross-dataset differences at the User-level were assessed using two-sided Mann-Whitney U tests, since the analyzed measures exhibit non-normal and heavy-tailed distributions. Statistical significance was complemented by Cliff's [8, 20] delta (δ) to quantify effect magnitude. Following conventional thresholds, $|\delta| < 0.147$ was interpreted as negligible, $|\delta| < 0.330$ as small, $|\delta| < 0.474$ as medium, and larger values as large.

5 COMPARATIVE ANALYSIS OF INFORMATION DIFFUSION

This section presents a comparative analysis of information diffusion across the French and Brazilian datasets using the four-dimensional analytical framework introduced in Section 4. The analysis examines four complementary dimensions: repeated exposure (REI), social-source diversity (SEI), their relationship through exposure organization (RII), and temporal dynamics preceding adoption.

5.1 Comparative Overview

Adoption events in the Brazilian dataset are generally preceded by more observed exposures than in France, particularly for hashtags. Hashtag adoption is preceded, on average, by 169.29 introductions and 24.96 distinct adopting followees in Brazil, compared with 39.68 and 13.06 in France, respectively. The following sections examine these cross-dataset differences through the four analytical dimensions.

5.2 Frequency Dimension: Repeated Exposure Influence (REI)

Repeated Exposure Influence (REI) characterizes the frequency of repeated exposure preceding information adoption. Figures 1(a-b) show a consistent qualitative pattern across

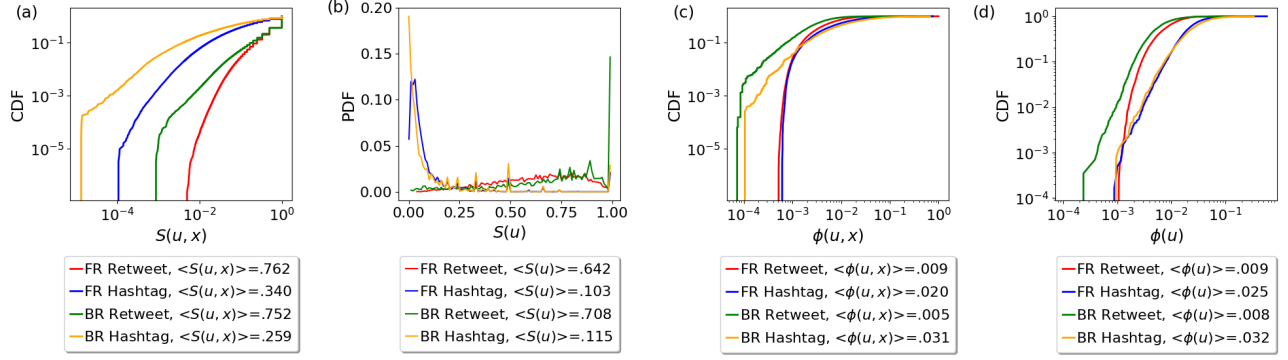


Figure 1: Distributions of Repeated Exposure Influence (REI) and Social Exposure Influence (SEI) for retweet and hashtag adoptions in the French and Brazilian datasets.

Source: Research data.

Interpretation: Panels (a) and (b) show the distributions of Repeated Exposure Influence (REI) at the *Item-level* and *User-level*, respectively. Panels (c) and (d) show the corresponding distributions of Social Exposure Influence (SEI). Each panel compares retweet and hashtag adoptions in the French (FR) and Brazilian (BR) datasets. Colors identify each dataset-content pair: FR retweets (red), FR hashtags (blue), BR retweets (green), and BR hashtags (orange). Mean values are reported in each legend.

Table 2: Topic-level summary of information diffusion metrics across hashtag groups in the French (FR) and Brazilian (BR) datasets.

Dataset	Topic group	Mechanisms		Entropy				RII				Influence	
		$\bar{S}(u, X_i)$	$\bar{\phi}(u, X_i)$	Avg. $\bar{R}(n)$	IQR	Max. n	% Adopt.	n	Median	IQR	p99	Avg. $I(u, X_i)$	% Adopt.
FR	Notre-Dame	0.111	0.016	0.588	0.093	20	67	1972	1.478	0.853	11.395	28.9	74
FR	Int. politics	0.153	0.013	0.632	0.065	13	58	943	1.500	1.000	8.000	31.5	76
FR	Tourism	0.216	0.010	0.480	0.264	16	69	1269	1.429	1.000	11.118	31.0	83
FR	French politics	0.116	0.016	0.543	0.095	17	61	1742	1.500	0.890	26.829	43.7	79
FR	EU elections	0.104	0.022	0.617	0.068	20	68	1438	1.500	0.947	7.053	30.9	77
FR	Entertainment	0.180	0.011	0.470	0.184	18	75	1290	1.500	1.463	12.436	25.7	80
FR	Protests	0.172	0.012	0.590	0.085	18	71	1685	1.364	0.817	6.883	22.3	76
FR	Climate	0.150	0.014	0.576	0.088	12	42	1784	1.800	1.122	10.387	53.1	77
FR	Gilets jaunes	0.110	0.019	0.613	0.054	20	61	1664	1.641	0.930	7.395	60.9	81
FR	Sports	0.130	0.017	0.552	0.142	12	62	924	1.417	0.983	10.383	23.1	77
BR	Int. entertainment	0.136	0.028	0.431	0.212	13	51	1385	2.084	3.545	40.530	60.0	81
BR	Brazilian COVID-19	0.121	0.028	0.457	0.246	18	55	1249	2.217	2.695	37.849	122.5	84
BR	Bolsonaro supporters	0.271	0.054	0.561	0.147	10	40	530	2.000	2.469	21.770	193.3	81
BR	Int. COVID-19	0.136	0.021	0.534	0.171	17	61	837	1.699	1.189	24.746	51.6	81
BR	Brazilian entertainment	0.082	0.023	0.349	0.246	20	50	1470	3.059	5.131	84.245	194.2	82
BR	Protests	0.180	0.025	0.543	0.158	8	54	1578	1.525	1.821	25.943	45.0	86
BR	Brazilian politics	0.211	0.042	0.547	0.173	12	59	779	1.833	2.133	66.819	74.0	81

Source: Research data.

Interpretation: Topic-level summary of REI, SEI, entropy, RII, influence-sequence length, and adoption rates across French and Brazilian hashtag groups.

both datasets: retweet adoption is preceded by fewer exposures than hashtag adoption. However, important quantitative differences are observed between the two datasets.

At the Item-level, the difference is driven mainly by hashtag adoption. The average number of prior retweet exposures increases only modestly from 2.69 in France to 3.27 in Brazil, whereas the average number of prior hashtag introductions increases from 39.68 to 169.29. This contrast shows that the cross-dataset difference in repeated exposure is substantially more pronounced for hashtags than for retweets.

The User-level results (Figure 1b) preserve the same qualitative distinction between content types. User-level REI distributions remain heterogeneous in both datasets, with larger cross-dataset differences observed for hashtags than for

retweets. Thus, the Item-level differences are also reflected in the User-level distributions rather than being restricted to isolated adoption events. The User-level cross-dataset differences are statistically detectable for both content types (Mann-Whitney U , $p < 0.001$), but their effect magnitudes are limited. For retweets, the difference corresponds to a small effect ($\delta = 0.222$), with median REI values of 0.777 in Brazil and 0.667 in France. For hashtags, the effect is negligible ($\delta = -0.125$), with medians of 0.038 and 0.050, respectively.

At the Topic-level (Table 2), repeated exposure also varies across thematic domains. France exhibits a narrower observed range of Topic-level REI values (0.104–0.216), whereas Brazil presents a broader observed range (0.082–0.271). This broader observed range is illustrated by *Bolsonaro Supporters* (0.271)

and *Brazilian Politics* (0.211), indicating substantial topic-level heterogeneity in the Brazilian dataset.

Overall, REI reveals heterogeneous patterns across content types and datasets. Although the User-level cross-dataset differences are statistically detectable, their effect sizes are small for retweets and negligible for hashtags.

5.3 Diversity Dimension: Social Exposure Influence (SEI)

Social Exposure Influence (SEI) characterizes the diversity of social exposure preceding information adoption. Figures 1(c–d) show that adoption is preceded by exposure from multiple social sources in both datasets, with hashtags generally associated with broader social exposure than retweets.

At the Item-level, hashtag adoption is preceded by a larger number of distinct adopting followees in both datasets. This difference is more pronounced in Brazil, where the average number of distinct followees preceding hashtag adoption nearly doubles compared with France (24.96 vs. 13.06), indicating broader observed social exposure preceding hashtag adoption in the Brazilian dataset.

The User-level distributions (Figure 1d) preserve the same qualitative distinction between content types. In both datasets, hashtag adoption is associated with exposure from a larger fraction of distinct followees than retweet adoption, although the distributions differ in magnitude across datasets. At the User-level, cross-dataset differences are also statistically detectable for both content types ($p < 0.001$), although effect sizes remain limited. Retweets exhibit a small effect ($\delta = -0.245$), with median SEI values of 0.0049 in Brazil and 0.0071 in France, whereas hashtags exhibit a negligible effect ($\delta = 0.102$), with medians of 0.0224 and 0.0195, respectively.

At the Topic-level (Table 2), SEI also varies across thematic domains and datasets. The observed Topic-level SEI values range from 0.010 to 0.022 in France and from 0.021 to 0.054 in Brazil. *Bolsonaro Supporters* (0.054) and *Brazilian Politics* (0.042) exhibit the highest Topic-level SEI values, while the remaining Brazilian topic groups range from 0.021 to 0.028.

Overall, SEI exhibits statistically detectable but limited cross-dataset differences at the User-level, with a small effect for retweets and a negligible effect for hashtags. These results reinforce the heterogeneous character of social-source diversity across content types rather than indicating a uniformly stronger pattern in either dataset.

5.4 Interaction Dimension: Reinforced Influence Index (RII)

The previous analyses characterized two complementary properties of exposure preceding adoption: exposure frequency through REI and social-source diversity through SEI. The Interaction Dimension examines the relationship between these properties by characterizing exposure volume relative to the number of distinct social sources. RII operationalizes this relationship as the average number of observed exposure events per distinct adopting followee and is therefore interpreted as a diagnostic of exposure organization.

At the Item-level, RII distributions differ between the two datasets. French adoption events exhibit a median RII of 1.50 (IQR = 1.33), whereas Brazilian events present a higher median of 1.93 and greater variability (IQR = 2.80), indicating higher and more heterogeneous exposure-per-source ratios in the Brazilian dataset.

At the User-level, Brazilian users exhibit a higher median RII than French users (3.207 versus 2.312) and a larger IQR (3.990 versus 1.417). The cross-dataset difference is statistically detectable ($p < 0.001$), with a small effect size ($\delta = 0.217$). For the same hashtag User-level populations, the corresponding effects for REI and SEI are negligible ($\delta = -0.125$ and $\delta = 0.102$, respectively). Thus, RII exhibits a more pronounced cross-dataset difference than either dimension considered separately, although its effect remains small.

At the Topic-level (Table 2), RII also reveals heterogeneity in exposure organization. French topic medians range from 1.36 to 1.80, whereas Brazilian topics span a broader interval from 1.52 to 3.06. For example, *Brazilian Entertainment* reaches a median RII of 3.06 (IQR = 5.13), while the highest French median, observed for *Climate*, is 1.80 (IQR = 1.12).

Overall, RII complements REI and SEI by characterizing exposure volume relative to the distinct social sources from which exposures originate. Differences are observed across all three aggregation levels, with the Brazilian dataset generally exhibiting higher exposure-per-source ratios and greater variability. At the User-level, the statistical comparison further shows that the cross-dataset difference is small but more pronounced for RII than for REI or SEI considered separately.

5.5 Temporal Dimension

The previous sections showed that repeated and diversified social exposure exhibit distinct patterns across topics and datasets. The temporal dimension extends this analysis by examining how these exposure patterns are organized before adoption through influence sequences, entropy, and last-event probabilities.

Influence sequences reveal substantially longer exposure histories in Brazil. At the Topic-level, average sequence lengths reach 194.2 for *Brazilian Entertainment*, 193.3 for *Bolsonaro Supporters*, and 122.5 for *Brazilian COVID-19*, whereas the longest French sequences are observed for *Gilets Jaunes* (60.9) and *Climate* (53.1), with most remaining below 45 exposures (Table 2). These results indicate substantially longer observed reinforcement sequences preceding adoption in the Brazilian dataset.

Entropy analysis further reveals differences in the organization of these sequences (Figure 2). France exhibits higher average entropy ($\bar{R}n = 0.47$) and a later entropy peak ($n = 30$), whereas Brazil presents lower average entropy ($\bar{R}n = 0.37$) and an earlier peak ($n = 15$). French sequences therefore maintain a more balanced combination of new and repeated exposures over longer sequence lengths. Brazilian sequences become less balanced at shorter sequence lengths and exhibit

greater variability at longer lengths, indicating more heterogeneous trajectories in the organization of new and repeated exposures.

Last-event probabilities provide complementary evidence about the final stage of adoption. In both datasets, the last exposure preceding adoption is more frequently a repeated exposure ($P_R = 0.76$ in France and 0.64 in Brazil), showing a predominance of repeated over new-source exposures at the final observed step. However, Brazilian sequences exhibit a higher probability than French sequences that this final exposure originates from a newly adopting followee ($P_N = 0.36$ versus 0.24).

From a temporal perspective, these results show that cross-dataset differences extend beyond the length of exposure histories to their organization before adoption. French sequences exhibit a more balanced organization of new and repeated exposures over longer sequence lengths, whereas Brazilian sequences are substantially longer and exhibit greater heterogeneity at longer lengths, while also presenting a higher probability of a new social source at the final observed exposure before adoption.

5.6 Cross-Context Synthesis

The comparative analysis across the four analytical dimensions reveals both recurring patterns and substantial heterogeneity in information diffusion across Brazil and France datasets. Across the French and Brazilian datasets, repeated and diversified social exposure are consistently observed preceding adoption, with retweets generally preceded by fewer exposures than hashtags and substantial heterogeneity observed across users, topics, and datasets.

Differences between the datasets are observed in the magnitude, joint organization, and temporal structure of reinforcement preceding adoption. In the Brazilian dataset, hashtag adoption is preceded by substantially more repeated exposures, while RII distributions are broader and influence sequences are considerably longer. French sequences exhibit higher average entropy and a later entropy peak, indicating a more balanced temporal organization of new and repeated exposures over longer sequence lengths.

These findings show that repeated and diversified social exposure are observed in both empirical contexts, while their magnitude, joint organization, and temporal structure differ across topics and datasets. Rather than reducing these differences to a single reinforcement mechanism, the four-dimensional framework shows that comparative diffusion patterns emerge across multiple complementary dimensions of exposure preceding adoption.

6 DISCUSSION

Previous studies have predominantly characterized information diffusion through global network statistics, probabilistic models, or aggregate adoption patterns. Building upon a previously proposed ego-network framework, this work provides a complementary user-centered perspective by integrating four

analytical dimensions: Frequency (REI), Diversity (SEI), Interaction (RII), and Temporal Dynamics. Together, these dimensions characterize repeated exposure, social-source diversity, exposure volume relative to distinct adopting followees, and the temporal organization of reinforcement preceding adoption.

The comparative analysis reveals both recurring patterns and substantial heterogeneity. Retweet adoption is generally preceded by fewer exposures than hashtag adoption in both datasets, while REI and SEI vary across topics and datasets. These findings support a multidimensional interpretation of reinforcement, in which repetition and social diversity are observed across both empirical contexts but vary across content types, topics, and datasets. Because the datasets also differ in geographic, temporal, topical, and event-related characteristics, these differences should be interpreted comparatively rather than causally.

RII complements these separate perspectives by relating exposure volume to the number of distinct adopting followees. At the User-level, RII exhibits a small cross-dataset effect for hashtag adoption, whereas the corresponding REI and SEI effects are negligible. Thus, exposure-per-source organization provides a complementary comparative perspective that is less apparent when frequency and social-source diversity are examined separately. Topic-level results further reveal heterogeneity in this organization across thematic groups.

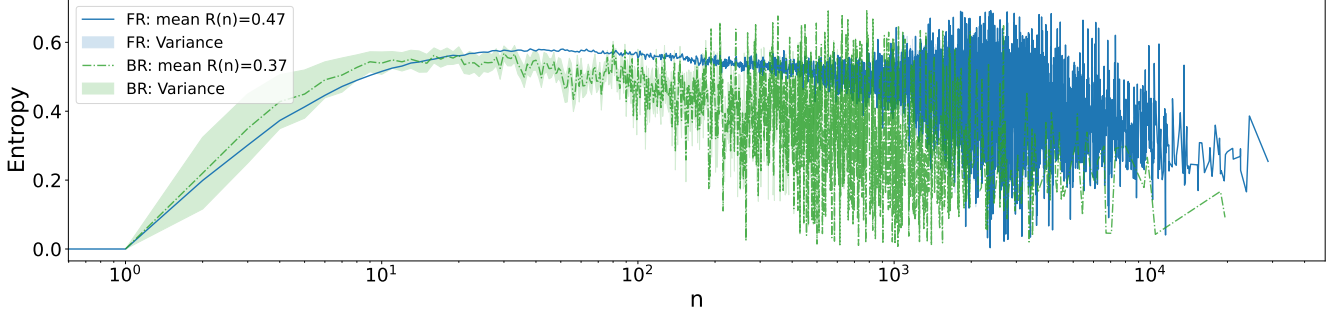
Temporal analyses extend these findings by showing differences in how reinforcement unfolds before adoption. Brazilian sequences are substantially longer and exhibit an earlier entropy peak, lower average entropy, and greater variability at longer sequence lengths, indicating more heterogeneous trajectories in how new and repeated social exposures are organized before adoption. Although repeated exposure is the most probable final event in both datasets, Brazilian sequences exhibit a higher probability than French sequences that the final observed exposure originates from a new social source ($P_N = 0.36$ vs. 0.24). Together, these results indicate that cross-dataset differences extend beyond exposure magnitude to the temporal organization of new and repeated exposures preceding adoption.

Overall, integrating Frequency, Diversity, Interaction, and Temporal Dynamics provides an interpretable multidimensional characterization of information diffusion and supports systematic comparisons across empirical contexts.

7 CONCLUSION

This work presented a comparative analysis of information diffusion across two large-scale Twitter (now X) datasets representing distinct contexts: the 2019 European Parliament election in France and the COVID-19 pandemic in Brazil. Building upon our previous ego-network representation of information adoption, we extended this analytical basis into four complementary dimensions: frequency, diversity, interaction, and temporal dynamics.

The comparative analyses revealed both recurring patterns and substantial heterogeneity across the four analytical

Figure 2: Entropy for Item-level.

Source: Research data.

Interpretation: Average sequence entropy Rn by sequence length; shaded areas represent one standard deviation.

dimensions. Repeated and diversified social exposure exhibited distinct patterns across content types, users, topics, and datasets, with retweet adoption generally preceded by fewer exposures than hashtag adoption. RII further revealed differences in exposure organization across aggregation levels, complementing the comparatively limited User-level cross-dataset differences observed for REI and SEI. Together, these results show that cross-dataset differences extend across multiple complementary properties of exposure preceding adoption rather than being restricted to a single analytical dimension.

Beyond the empirical findings, this work contributes a unified analytical framework for studying information diffusion from individual ego-networks. The framework integrates four complementary dimensions and introduces the Reinforced Influence Index (RII) to operationalize the Interaction Dimension as a diagnostic of exposure organization, relating observed exposure volume to the number of distinct adopting followees. In addition, this work presents the first analysis of the Brazilian COVID-19 ego-network dataset and revisits the previously studied French dataset under the same analytical perspective, enabling their first unified comparison.

Although the proposed framework is independent of a particular platform or diffusion event, the empirical evaluation covers Twitter data from two specific contexts. Because the datasets differ simultaneously in geographic, temporal, topical, and event-related characteristics, the observed differences should be interpreted comparatively and cannot be attributed to any single contextual factor. Nevertheless, because the four dimensions are defined from user-level exposure properties rather than platform-specific features, the framework provides a general basis for investigating information adoption in other online social systems.

Future work includes validating the framework on additional social platforms and diffusion scenarios, investigating whether the recurring and heterogeneous exposure patterns identified across the four dimensions generalize to other empirical contexts, and developing predictive models that integrate these dimensions to improve the understanding and anticipation of information adoption.

8 FINAL DECLARATIONS

8.1 Data Availability

The datasets analyzed in this study were constructed exclusively from publicly accessible information obtained through the Twitter Academic Application Programming Interface (API) [23], in accordance with the platform’s Developer Policy and Academic Research Terms of Service in effect at the time of data collection. The collected data comprise publicly available user profile information, tweets, and follow relationships required to reconstruct ego-networks and information adoption events.

To promote transparency and reproducibility, the derived datasets, analytical metrics, and source code supporting the reported analyses will be made publicly available upon publication through a dedicated online repository, subject to the Twitter terms of service and applicable data-sharing policies. Any released data will consist exclusively of processed, derived, or aggregated information and will not include unnecessary user identifiers.

8.2 Ethical and Legal Considerations

This study relies exclusively on publicly accessible social media data and involves no direct interaction with human participants. No private communications, restricted-access information, or sensitive personal data were collected or processed, and all results are reported in aggregate form without identifying individual users.

The research complies with the Brazilian General Data Protection Law (Lei Geral de Proteção de Dados – LGPD), the European Union General Data Protection Regulation (GDPR), and Resolution No. 510/2016 of the Brazilian National Health Council. Data collection was limited to publicly available information necessary to reconstruct information diffusion processes. The analyses focus on aggregate behavioral patterns and involve no profiling or inference of sensitive personal attributes. Therefore, the risk of re-identification or individual harm is considered minimal.

8.3 Research Transparency

The methodological procedures, including data collection criteria, ego-network reconstruction, metric definitions, and analytical workflow, are described in the corresponding sections of this paper. Additional details are available in the associated publications. Limitations related to data availability, platform constraints, and Twitter data should be considered when interpreting and reproducing the analyses.

8.4 Use of Artificial Intelligence Tools

Generative Artificial Intelligence (GenAI) tools were used exclusively to support drafting assistance, language revision, grammar correction, and improvements in textual clarity during manuscript preparation. All scientific decisions, including the research design, data collection procedures, metric definitions, analyses, interpretation of results, and conclusions, were developed, validated, and approved by the authors.

9 ACKNOWLEDGMENTS

This work was supported by CNPq (Grant #409640/2025-5), FAPEMIG, and Instituto Nacional de Ciência e Tecnologia em Inteligência Artificial Responsável para Linguística Computacional, Tratamento e Disseminação de Informação (INCT-TILD-IAR – Grant #408490/2024-1). The authors thank the Pontifícia Universidade Católica de Minas Gerais – PUC Minas (FIP #2026-35242) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES – (Grant PROAP 88887.842889/2023-00 / PUC Minas - INFORMÁTICA and Finance Code 001).

REFERENCES

- [1] Elsa Andres, Gergely Ódor, Iacopo Iacopini, and Márton Karsai. 2025. Distinguishing mechanisms of social contagion from local network view. *npj Complexity* 2, 1 (2025), 8.
- [2] Pedro Bento, Arthur Buzelin, Yan Aquino, Isis Carvalho, Pedro Dutenhefner, Lucas Dayrell, Caio Santana, Victoria Estanislau, Gisele L Pappa, Debora Miranda, et al. 2024. Impacto da Pandemia na Discussão sobre Saúde Mental: O Caso do Discord no Brasil. In *Brazilian Symposium on Multimedia and the Web (WebMedia)*. SBC, 179–187.
- [3] Ronald S. Burt. 1992. *Structural Holes: The Social Structure of Competition*. Harvard University Press, Cambridge, MA.
- [4] Josemar Alves Caetano, Humberto Torres Marques-Neto, and Virgílio Almeida. 2025. Characterizing Collective Attention on Online Chats: A Three-Pronged Approach. In *International Conference on Web Information Systems Engineering*. Springer, 56–66.
- [5] Damon Centola. 2010. The spread of behavior in an online social network experiment. *Science* 329, 5996 (2010), 1194–1197. doi:10.1126/science.1185231
- [6] Damon Centola. 2018. *How behavior spreads: The science of complex contagions*. Vol. 3. Princeton University Press Princeton, NJ.
- [7] Damon Centola, Víctor M Eguíluz, and Michael W Macy. 2007. Cascade dynamics of complex propagation. *Physica A: Statistical Mechanics and its Applications* 374, 1 (2007), 449–456.
- [8] Norman Cliff. 1993. Dominance Statistics: Ordinal Analyses to Answer Ordinal Questions. *Psychological Bulletin* 114, 3 (1993), 494–509. doi:10.1037/0033-2909.114.3.494
- [9] Jaqueline Faria De Oliveira, Humberto Torres Marques-Neto, and Márton Karsai. 2020. Information adoption via repeated or diversified social influence on Twitter. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 237–241.
- [10] Jaqueline Faria de Oliveira, Humberto Torres Marques-Neto, and Márton Karsai. 2022. Measuring the effects of repeated and diversified influence mechanism for information adoption on Twitter. *Social Network Analysis and Mining* 12, 1 (2022), 16.
- [11] Tobias Friedrich, Andreas Göbel, Michael Klodt, Martin S. Krejca, and Maximilian Pappik. 2024. From Market Saturation to Social Reinforcement: Understanding the Impact of Non-Linearity in Information Diffusion Models. *arXiv preprint arXiv:2401.10818* (2024). doi:10.48550/arXiv.2401.10818
- [12] Mark Granovetter. 1978. Threshold models of collective behavior. *American journal of sociology* 83, 6 (1978), 1420–1443.
- [13] Sara Heydari, Gerardo Iñiguez, János Kertész, and Jari Saramäki. 2024. Disentangling degree and tie strength heterogeneity in egocentric social networks. *EPJ Data Science* 13, 1 (2024), 76.
- [14] Tuan-Anh Hoang and Ee-Peng Lim. 2016. Tracking virality and susceptibility in social media. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, 1059–1068.
- [15] Gerardo Iñiguez, Sara Heydari, János Kertész, and Jari Saramäki. 2023. Universal patterns in egocentric communication networks. *Nature Communications* 14, 1 (2023), 5217.
- [16] Larissa Malagoli, Giovana Piorino, Carlos HG Ferreira, and Ana Paula Couto da Silva. 2024. Twitter and the 2022 Brazilian elections portrait: A network and content-driven analysis. In *Brazilian Symposium on Multimedia and the Web (WebMedia)*. SBC, 283–291.
- [17] Fanhui Meng, Jiarong Xie, Jiachen Sun, Cong Xu, Yutian Zeng, Xiangrong Wang, Tao Jia, Shuhong Huang, Youjin Deng, and Yanqing Hu. 2025. Spreading dynamics of information on online social networks. *Proceedings of the National Academy of Sciences* (2025). doi:10.1073/pnas.2410227122
- [18] Pedro Ramaciotti Morales and Jean-Philippe Cointet. 2021. Online Social Movements Dynamics in Ideological Spaces. In *7th International Conference on Computational Social Science IC2S2*.
- [19] Lucas Santos De Oliveira, Wesley Costa, Pedro Olmo Stancioli Vaz De Melo, and Fabrício Benevenuto. 2022. How politicians communicate in social media: A cross-platform study. In *Proceedings of the Brazilian Symposium on Multimedia and the Web*. 75–83.
- [20] Jeanine Romano, Jeffrey D. Kromrey, Jesse Coraggio, and Jeff Skowronek. 2006. Appropriate Statistics for Ordinal Level Data: Should We Really Be Using t-Test and Cohen's d for Evaluating Group Differences on the NSSE and Other Surveys?. In *Annual Meeting of the Florida Association of Institutional Research*.
- [21] Zhongbao Ruan, Gerardo Iñiguez, Márton Karsai, and János Kertész. 2015. The impact of heterogeneous thresholds on social contagion with multiple initiators. *PLOS ONE* 10, 9 (2015), e0137004. doi:10.1371/journal.pone.0137004
- [22] Bogdan State and Lada A. Adamic. 2014. The diffusion of support in an online social movement: Evidence from the adoption of equal-sign profile pictures. *PLOS ONE* 9, 3 (2014), e90036. doi:10.1371/journal.pone.0090036
- [23] Inc Twitter. 2019. *Twitter Developer*. <https://developer.twitter.com>
- [24] Samuel Unicomb, Gerardo Iñiguez, János Kertész, and Márton Karsai. 2019. Reentrant phase transitions in threshold driven contagion on multiplex networks. *Phys. Rev. E* 100 (Oct 2019), 040301. Issue 4. doi:10.1103/PhysRevE.100.040301
- [25] Samuel Unicomb, Gerardo Iñiguez, and Márton Karsai. 2018. Threshold driven contagion on weighted networks. *Scientific reports* 8, 1 (2018), 3094.
- [26] Fei Wang, Zhenfang Zhu, Peiyu Liu, and Peipei Wang. 2019. Influence maximization in social network considering memory effect and social reinforcement effect. *Future internet* 11, 4 (2019), 95.
- [27] Duncan J Watts. 2002. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences* 99, 9 (2002), 5766–5771. doi:10.1073/pnas.082090499
- [28] Z. Zhang, H. Xu, and J. Zhang. 2021. Social media-based disaster research: Development, trends, and obstacles. *International Journal of Disaster Risk Reduction* 56 (2021), 102142. doi:10.1016/j.ijdrr.2021.102142
- [29] Muhua Zheng, Linyuan Lü, and Ming Zhao. 2013. Spreading in online social networks: The role of social reinforcement. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 88, 1 (2013), 012818.
- [30] Tao Zhou, Zhi-Dan Zhao, and Zhong-Yuan Lu. 2015. Impact of repeated exposures on information spreading in social networks. *PLoS ONE* 10, 10 (2015), e0140556. doi:10.1371/journal.pone.0140556