BG Pulse Journal USA
ISSN:1947-2536 http://www.bgpulseusa.com/index.php/bg

Who Breaks the News? Measuring Concentration, Latency, and Diffusion Structure in Uzbekistan's Telegram News Ecosystem

Evidence from Channel Citation Metrics and Story Tracing

Author's: **Asomiddin Asomiddinov**
[ORCID ID: 0009-0004-4582-4684,
asomiddinabroad@gmail.com]
Mo'minjon Muydinjonov [ORCID ID: 0009-0007-2719-2953,
mominpolo2@gmail.com]
Azizxon Rasulov [ORCID ID: 0009-0000-7362-7216,
azizkhandimorom@gmail.com]
Ramziddin Todjiddinov [ORCID ID: 0009-0003-6118-5870,
rtadjiddinov@gmail.com]
Target International School, Tashkent, Uzbekistan
doi:https://doi.org/10.5281/zenodo.23525620

Abstract: Telegram is the main news infrastructure in Uzbekistan, a country of 37 million people with an estimated 25 million Telegram users. Because the platform makes republication effortless and attribution optional, a few newsrooms may supply the stories that hundreds of channels redistribute. This paper asks how centralised the flow of information in this ecosystem is and examines three dimensions: concentration, latency and diffusion structure. It combines two sources of real data. The first is a cross-section of the 100 largest Uzbek news channels in the TGStat catalogue, covering subscribers, average post reach and the citation index, a measure of how often a channel is cited and reposted by other channels. The second is a manual trace of 25 story families (55 posts) identified in the public web previews of 15 leading channels (264 posts). Citations are markedly more concentrated than audiences: the Gini coefficient is 0.61 for the citation index against 0.46 for subscribers, and the bottom half of channels receives only 8.2% of citations but holds 19.8% of subscribers. Concentration is nonetheless moderate in absolute terms (CR_{10} = 38.9%, HHI = 259), and the largest channel by subscribers ranks only 48th by citations. In the traced families, the first republication followed the root after a median of 20 minutes (all republications: median 38 minutes, IQR 10–124). None of the 264 posts was a native forward of another news channel, so republication took place entirely through unattributed copying. Rather than a single-hub star, the observed structure is one of reciprocal copying among peer channels, with official sources feeding professional media. Root posts reached a larger share of their channel's audience than copies in 64% of families (Wilcoxon p = 0.06). We discuss the implications for the economics of originality and for attribution norms, and outline an embedding-based pipeline for scaling the analysis.

Keywords: Telegram; news diffusion; media concentration; citation index; copy-and-paste journalism; network analysis; media economics; Uzbekistan

1. Introduction

In much of the post-Soviet space, the messenger Telegram has become core news infrastructure, and Uzbekistan is an extreme case. In October 2025 the country had 33.1 million

internet users, 89 per cent of its population of 37.2 million (Kepios, 2026).[1] The Telegram audience is estimated at nearly 25 million people (Donaev, 2024), and by 2023 the number of channels registered to Uzbek users exceeded 123,000, second only to Russia (Kun.uz, 2023). TGStat now lists more than 7,000 Uzbek channels in its news category alone (TGStat, 2026). Newsrooms, ministries, regional administrations and anonymous aggregators all publish through Telegram channels,[2] and survey evidence identifies Telegram as the leading messenger for news in the country (CABAR.asia, 2020).

This matters economically. News is costly to produce, but once published its information behaves like a public good,[3] and on Telegram the cost of reproduction is close to zero. Cagé, Hervé and Viaud (2020)[4] showed that in French online media only about one third of content was original and a quarter of stories were reproduced within four minutes, which erodes the returns to original reporting. Diffusion structure also matters for resilience. If most channels copy a few hubs (a star), errors at the hub spread everywhere at once. If stories pass from channel to channel (a chain), attribution and accuracy may degrade at each step.

No quantitative evidence exists on these questions for Uzbekistan. The research question of this paper is therefore: How centralised is the flow of information in the Uzbek Telegram news ecosystem? How many channels act as original sources, and how quickly and in what pattern do the others republish them? We examine three components: (RQ1) concentration, the share of source activity held by the top 5 and top 10 channels; (RQ2) latency, the time a story takes to travel from its first observed publisher to subsequent channels; and (RQ3) structure, whether republication is star-like, chain-like or reciprocal.

The paper makes three contributions. First, it provides the first systematic measurement of source concentration in Uzbek Telegram news, contrasting the distribution of citations with that of audiences. Second, it documents real republication sequences, their timing and the absence of attribution. Third, it links these patterns to media economics and proposes a scalable, embedding-based design for future large-scale replication. Section 2 reviews the literature, Section 3 describes the data and methods, Section 4 presents the results, Section 5 discusses them, and Section 6 concludes.

2. Literature Review and Hypotheses

Economics of copying. News markets combine high fixed costs of production with near-zero marginal costs of distribution (Hamilton, 2004). Competition can discipline bias (Gentzkow and Shapiro, 2008), but it also invites free-riding on rivals' reporting. Boczkowski (2010) documented how continuous monitoring of competitors produced homogeneous online news in Argentina, and British studies found heavy reliance on wire and public-relations copy, labelled "churnalism"

[1]Kepios / DataReportal. A digital-market research firm whose annual *Digital* reports combine data from the United Nations, GSMA Intelligence and platform advertising tools. Its figures rest partly on modelled data and are best read as approximate.

[2]Telegram channel. A one-to-many broadcast feed in the Telegram messenger. Only administrators post; subscribers read and may forward posts. Every post carries a public view counter and a timestamp. A post can be republished by the native *forward* function, which keeps a visible link to the source, or by copy-and-paste, which leaves no trace of the source.

[3]Public good. A good that is non-rival (one person's use does not reduce what is left for others) and non-excludable (non-payers cannot easily be prevented from using it), as formalised by Paul A. Samuelson (1954). Once published, the information in a news story has both properties.

[4]Julia Cagé is an economist at Sciences Po (Paris) working on the economics of the media. Her study with Nicolas Hervé and Marie-Luce Viaud (2020) is the main quantitative benchmark for copying between news outlets.

(Davies, 2008; Lewis, Williams and Franklin, 2008).[5] Cagé, Hervé and Viaud (2020) analysed 2.5 million French news documents and found that only 3.4 per cent of copying documents cited the outlet they copied, but that originality was rewarded with larger audiences.

Agenda setting and cascades. Intermedia agenda setting[6] describes how some outlets shape the coverage of others (McCombs and Shaw, 1972; Vliegenthart and Walgrave, 2008). Harder, Sevenans and Van Aelst (2017) found that high-frequency outlets, namely news websites and Twitter accounts, lead the agenda in the social media age. Information cascades[7] explain why unverified reports spread when prior publication is treated as a signal (Bikhchandani, Hirshleifer and Welch, 1992). Computational studies traced diffusion at scale (Leskovec, Backstrom and Kleinberg, 2009), inferred latent diffusion networks from timing (Gomez-Rodriguez, Leskovec and Krause, 2012), and distinguished broadcast from viral diffusion with structural virality[8] (Goel et al., 2016). Vosoughi, Roy and Aral (2018) showed that false news travels faster than true news.

Telegram research. Studies of Telegram have concentrated on extremism and politics. Examples include the Pushshift Telegram dataset (Baumgartner et al., 2020), forward networks of far-right channels (Urman and Katz, 2022) and Telegram as a political actor in post-Soviet states (Wijermars and Lokot, 2022). These studies infer links from native forwards. Copy-and-paste republication leaves no such trace and requires text-reuse detection, either n-gram matching (Nicholls, 2019) or multilingual sentence embeddings (Reimers and Gurevych, 2020; Feng et al., 2022; Chen et al., 2022). Table 1 summarises the closest empirical studies.

Table 1. Selected empirical studies on news copying and diffusion

Study	Setting and data	Key finding
Cagé, Hervé and Viaud (2020)	French online media, 2013; ~2.5 m documents	~33% of content original; 25% of stories copied in < 4 min; 3.4% of copies cite the source
Leskovec, Backstrom and Kleinberg (2009)	90 m news and blog articles	Regular news-cycle dynamics; mainstream media lead blogs by about 2.5 hours
Goel et al. (2016)	~1 bn Twitter diffusion events	Large cascades mostly broadcast-dominated (star-like)
Harder, Sevenans and Van Aelst (2017)	Belgian election news, multiple media	High-frequency online outlets set the agenda
Urman and Katz (2022)	Far-right Telegram channels	Forward links reveal hub-centred clusters

[5]Churnalism. A term attributed to BBC journalist Waseem Zakir and popularised by Nick Davies (2008) for journalism that recycles press releases, wire copy and rival outlets' stories with little independent reporting.

[6]Intermedia agenda setting. The influence some media outlets exert on which issues other outlets cover. It extends the agenda-setting theory of Maxwell McCombs and Donald Shaw (1972): the media do not tell people what to think, but what to think about.

[7]Information cascade. A situation in which agents rationally disregard their own information and imitate earlier agents (Bikhchandani, Hirshleifer and Welch, 1992). In news, outlets republish a report because others already have, treating prior publication as evidence of reliability.

[8]Structural virality. A measure proposed by Goel et al. (2016): the mean shortest-path distance between all pairs of nodes in a diffusion tree. Low values indicate broadcast (star-like) diffusion from one source; high values indicate multi-step, chain-like diffusion.

Study	Setting and data	Key finding
This study	100 Uzbek news channels + 25 traced story families	Citations more concentrated than audiences; unattributed reciprocal copying

Following this literature, we stated four expectations in advance: H1, source activity is highly concentrated (CR_{10} of citations above 50%); H2, republication is fast (median first-copy delay below 30 minutes); H3, diffusion is star-shaped around a few hubs; and H4, originators obtain more audience attention than copiers of the same story.

3. Data and Methods

Figure 1 summarises the digital context. Uzbek Telegram combines near-universal mobile access, a very large number of channels and almost no legal cost of copying, since the facts in news are not protected by copyright.[9] We use two complementary datasets (Table 2).

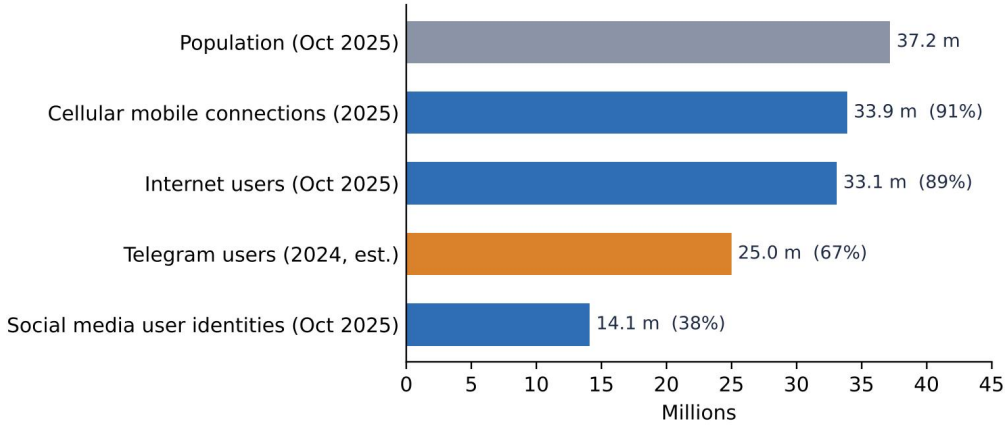


Figure 1. Digital access indicators for Uzbekistan (millions; share of population in brackets)[10]

Table 2. Data sources[11]

● Dataset	● Content	● Unit and size	● Used for
● A. TGStat news ranking	● Subscribers, average post reach, citation index of the top 100 channels in the category "News and Media", Uzbekistan	● Channel; n = 100 (robustness: n = 105)	● RQ1

[9] **News of the day.** Article 2(8) of the Berne Convention excludes "news of the day or … miscellaneous facts having the character of mere items of press information" from protection. Uzbekistan has been party to the Convention since 2005, so the facts in a story are free to reuse, while original wording and images are protected.

[10] Kepios (2026), data for October 2025; Telegram users from Donaev (2024), a press estimate not directly comparable with the Kepios series.

[11] TGStat (2026); Telegram public web previews. Both retrieved on 18 October 2026.

Dataset	Content	Unit and size	Used for
B. Story tracing	Posts visible in the public web previews of 15 leading channels; time, views, forward labels; manual matching into story families	Post; 264 posts → 25 families, 55 posts, 10 channels	RQ2, RQ3, H4

3.1 Dataset A: channel-level metrics

From the TGStat[12] catalogue for Uzbekistan, we collected the 100 largest channels in the category "News and Media" (all with about 100,000 subscribers or more), together with three indicators: subscribers, average post reach[13] and the citation index (CI).[14] The CI is our proxy for a channel's role as a source. A channel that others frequently mention or repost has a high CI, whatever the size of its own audience. For each indicator we compute the concentration ratios CR_5 and CR_{10},[15] the Herfindahl–Hirschman Index[16] and the Gini coefficient:[17]

$$s_c = x_c / \Sigma_j x_j, \quad CR_k = \Sigma_{c=1\ldots k}\, s_{(c)}, \quad HHI = \Sigma_c (100 \cdot s_c)^2 \qquad (1)$$

where x_c is the indicator for channel c and $s_{(c)}$ are the shares sorted in descending order. Uncertainty is quantified with 5,000-draw bootstrap intervals,[18] and the association between audience and source status with Spearman's ρ.[19] As a robustness check, we add five large news-producing channels that TGStat classifies under "Politics", including the President's press service, giving n = 105.

[12]TGStat. A commercial analytics service (tgstat.com) that tracks public Telegram channels and groups and publishes country and category rankings. Its regional catalogue for Uzbekistan (uz.tgstat.com) groups channels by thematic category; the category used here is *Yangiliklar va OAV* ("News and Media").

[13]Average post reach and ERR. Average post reach is the mean number of views a post accumulates (TGStat measures it over the first 24 hours). Dividing reach by the number of subscribers gives the engagement rate by reach (ERR), a common measure of how actively an audience reads a channel.

[14]Citation index (TGStat). A proprietary indicator of how often a channel is cited by other Telegram channels. It is built from mentions of the channel (links and @-references) and reposts of its posts in other channels, with greater weight given to citations from channels that are themselves highly cited, by analogy with citation indices in science. Paid promotional mentions are not fully separable from editorial citations, so the index is a proxy for, not a direct measure of, source status.

[15]Concentration ratio (CR_k). The combined share of the k largest units in a market; CR_4 and CR_8 are conventional in industrial organisation, while CR_5 and CR_{10} are used here to match the research question.

[16]Herfindahl–Hirschman Index (HHI). The sum of squared market shares in percentage points, ranging from near 0 (atomistic competition) to 10,000 (monopoly); named after Orris C. Herfindahl and Albert O. Hirschman (1964). The U.S. 2023 Merger Guidelines treat markets above 1,800 as highly concentrated.

[17]Gini coefficient and Lorenz curve. The Lorenz curve (Lorenz, 1905) plots the cumulative share of a quantity against the cumulative share of units ranked from smallest to largest; the Gini coefficient (Gini, 1912) is twice the area between this curve and the line of equality, ranging from 0 (equality) to 1 (one unit holds everything).

[18]Bootstrap confidence interval. A resampling method (Efron, 1979) in which the sample is redrawn with replacement many times (here 5,000) and the statistic is recomputed each time; the 2.5th and 97.5th percentiles of the resulting distribution form a 95% interval.

[19]Spearman's rank correlation (ρ). A non-parametric correlation computed on ranks rather than raw values. It ranges from −1 to 1 and is robust to the heavy-tailed distributions typical of audience data.

3.2 Dataset B: story tracing

For RQ2 and RQ3, we read the public web previews[20] of 15 leading channels, chosen from the top of the TGStat ranking to cover professional media (Kun.uz, Daryo, Qalampir), state sources (UzA, the presidential press service) and high-reach aggregators. Each preview showed between 7 and 20 posts, 264 posts in total, as served on 18 October 2026. Because the preview pages are cached, the visible posts belong to four different publication days. We label these S1 to S4 and date them from post content (S1: early August 2026; S2: early September; S3: mid-October; S4: the visit of a Chinese government delegation). For every post we recorded the publication time shown on the page (in the same time zone for all channels), the view count and any "Forwarded from" label.

Posts were matched manually into message families[21] when two or more channels on the same day reported the same specific event (for example, the same appointment, court verdict or proposal). Advertisements, quizzes and posts without text were excluded. This yielded 25 families comprising 55 posts from 10 channels: 20 families with two posts and 5 with three. For each family, the earliest post is the root, and each later post has a root latency $L = t - t_{root}$ in minutes. We summarise latencies, map root → copier links into a directed network, and compare audience uptake (views divided by subscribers) between roots and copies with the Wilcoxon signed-rank test.[22]

4. Results

4.1 RQ1: Source status is more concentrated than audience

Table 3 reports concentration indices for the three indicators. All three are unequally distributed, but citations markedly more so. The Gini coefficient for the citation index is 0.61 (95% CI 0.55–0.66), compared with 0.46 (0.40–0.50) for subscribers and 0.49 for reach; the intervals for citations and subscribers do not overlap. The difference lies mainly in the lower tail. The 50 smallest channels by each measure hold 19.8% of subscribers but only 8.2% of citations, and 7 channels with 100,000 or more subscribers have a citation index below 10, that is, they are almost never cited. The top 20 channels account for 64.2% of citations (Figure 2).

Table 3. Concentration of audience and source status among the 100 largest Uzbek news channels

Indicator	CR_5	CR_{10} (95% CI)	Top-20 share	HHI (95% CI)	Gini (95% CI)
Subscribers	20.5%	36.0% (30–40)	55.4%	207 (175–243)	0.46 (0.40–0.50)
Average post reach	21.2%	32.9%	51.7%	214	0.49

[20]Telegram web preview. Every public channel has a read-only web page at the address t.me/s/<username> that displays its most recent posts with their text, view counts, time of publication and any "Forwarded from" label, without requiring a Telegram account.

[21]Message family. The set of posts, published by different channels, that report the same specific news event. The earliest post in a family is its *root* (first observed publisher); later posts are *republications*.

[22]Wilcoxon signed-rank test. A non-parametric test of whether paired differences are centred on zero (Wilcoxon, 1945). It is preferred to the paired t-test for small samples with skewed distributions.

Indicator	CR_5	CR_{10} (95% CI)	Top-20 share	HHI (95% CI)	Gini (95% CI)
Citation index	22.6%	38.9% (32–46)	64.2%	259 (212–317)	0.61 (0.55–0.66)
Robustness, n = 105: citation index	24.8%	39.3%	62.1%	268	0.62
Robustness, n = 105: subscribers	21.1%	35.1%	55.5%	210	0.49

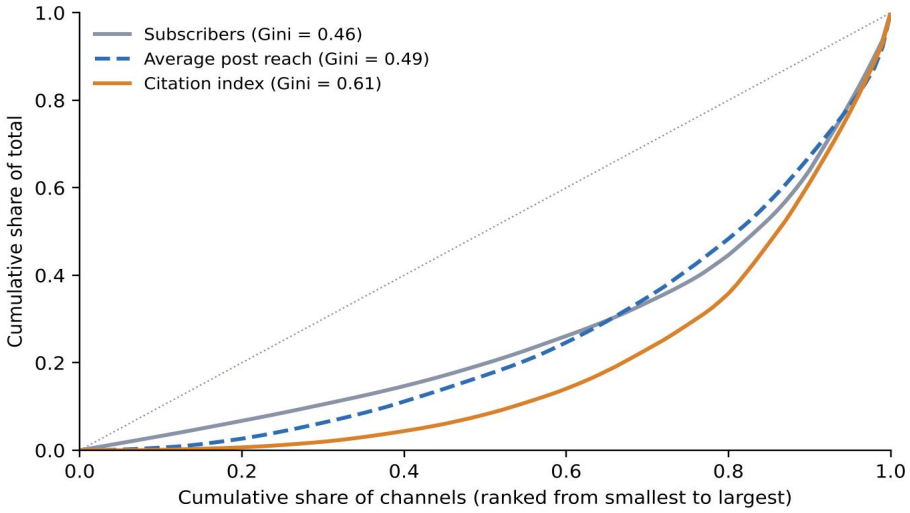


Figure 2. Lorenz curves of subscribers, average post reach and citation index across the 100 largest Uzbek news channels[23]

In absolute terms, however, concentration is moderate. CR_{10} for citations is 38.9% (95% CI 32–46%) and the HHI is 259, far below the thresholds that competition authorities treat as concentrated. H1 is therefore not supported in its strong form: no small cartel of channels supplies the majority of citations. Source status is distributed across about 20 channels, and below them is a long tail of channels that are almost never cited.

[23] Source: author's calculations based on TGStat (2026).

Source status and audience are related, but far from identical (Figure 3). Spearman's ρ between subscribers and citation index is 0.60 ($p < 0.001$). Table 4 lists the ten most-cited channels. Kun.uz stands out, receiving 7.0% of all citations, 1.6 times the next channel. The largest channel by subscribers, Yaxshi Yomon (1.93 million), has a citation index of only 339 and ranks 48th by citations. Conversely, @tezkor24official ranks 41st by subscribers but 6th by citations. The ten largest channels by audience together receive only 31.9% of citations. Reaching a large audience and acting as a source are, to a substantial degree, separate roles.

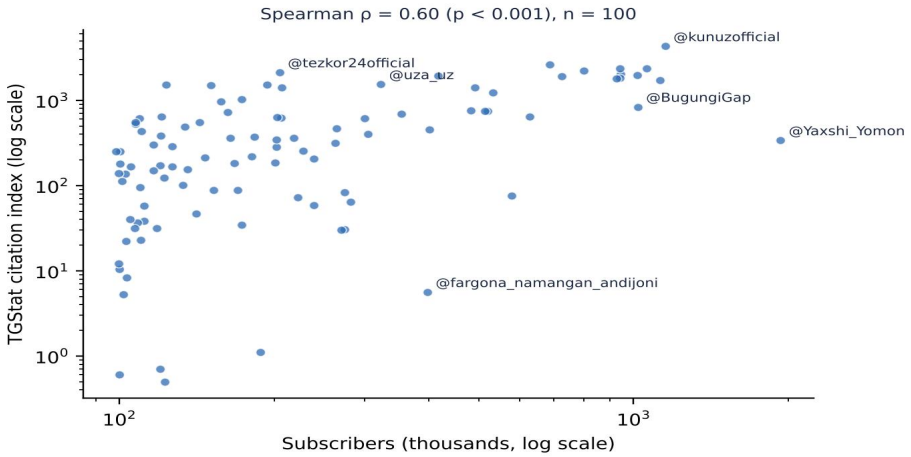


Figure 3. Subscribers and citation index of the 100 largest Uzbek news channels

Source: author's calculations based on TGStat (2026). Both axes logarithmic; one channel with a citation index of 0 is plotted at 0.5.

Table 4. The ten most-cited Uzbek news channels

CI rank	Channel	Subscribers	Rank by subscribers	Avg. post reach	Citation index	Share of all citations
1	@kunuzofficial	1,152,303	2	275,400	4,300	7.0%
2	@uzbekistanofficial	688,005	13	46,400	2,616	4.3%
3	@shopirlar	1,061,599	4	90,300	2,376	3.9%
4	@uzbekistan24	941,162	8	74,800	2,365	3.9%
5	@qoraxabar	799,826	11	86,600	2,216	3.6%
6	@tezkor24official	204,880	41	38,000	2,126	3.5%
7	@patrul_uz	942,383	7	81,700	2,042	3.3%
8	@TYUZBEK	1,017,166	6	112,500	1,961	3.2%
9	@UnchaMuncha	416,850	21	40,300	1,930	3.1%
10	@Jx_uzb	724,170	12	38,400	1,914	3.1%

Source: TGStat (2026); shares are author's calculations over the 100 channels.

4.2 RQ2: Republication latency

Figure 4 shows the timing of all 25 traced families, and Table 5 summarises the latencies. The first republication followed the root after a median of 20 minutes. Across all 30 republications, the median root latency was 38 minutes (IQR 9.5–124; 90th percentile 296). 40% of republications

appeared within 15 minutes and 63% within an hour, but the distribution has a long right tail of up to 622 minutes. The fastest copy came after 2 minutes. Only one family (4%) was first copied within 4 minutes, compared with 25% in the French benchmark (Cagé, Hervé and Viaud, 2020).

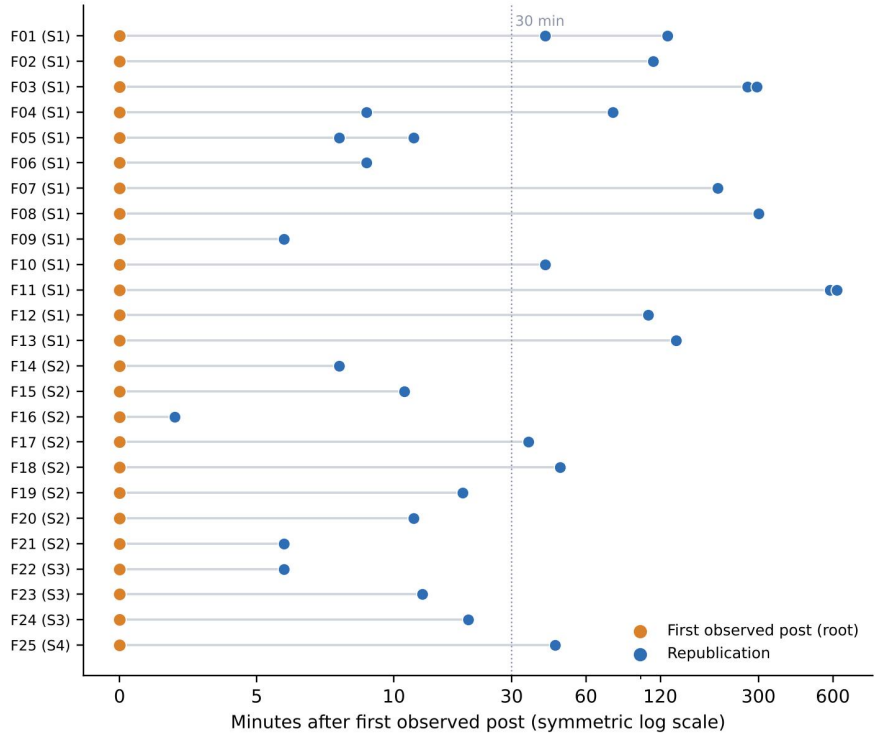


Figure 4. Timing of republication in 25 traced story families[24]

Latency differs sharply between pairs of channels. Between @patrul_uz and @qoraxabar (S2), all eight stories were republished within 47 minutes (median 11.5), and the two channels published near-identical sequences of stories within minutes of each other. The same holds for @uzbekistan24 and @TezkorOzbekiston (S3; median 13 minutes). Among the four large channels in S1, the median was 110 minutes, and several morning stories were republished only in the afternoon. **H2 is partly supported**: the median first-copy delay is below 30 minutes, but the overall distribution is slower and more heterogeneous than the French benchmark.

Table 5. Republication latency and audience uptake in traced families[25]

Measure	Value
Story families / posts / channels	25 / 55 / 10
Median first-copy latency (min)	20
Median latency, all republications (IQR) (min)	38 (9.5–124)
Share of republications within 15 / 30 / 60 min	40% / 47% / 63%

[24] Source: author's coding of Telegram public web previews (retrieved 18 October 2026). Orange: first observed post (root); blue: republications. S1–S4 denote publication days.

[25] Uptake = views ÷ subscribers × 100; subscriber counts from TGStat (2026).

Measure	Value
Median latency by day: S1 / S2 / S3 / S4 (min)	109.5 / 11.5 / 13 / 45
Native forwards from other news channels (of 264 posts)	0
Median views/subscribers: roots vs. copies	6.0% vs. 5.8%
Families in which the root had higher uptake	64% (Wilcoxon p = 0.06)

4.3 RQ3: Reciprocal copying rather than a single hub

Figure 5 maps the root → copier links. Three features stand out. First, **no channel acts as a** sole hub. In S1, roots were spread across all four channels (@BugungiGap 6, @Yaxshi_Yomon 4, @uzbekistanofficial 2, @TYXUZBEK 1), and in S2 @patrul_uz published first five times and @qoraxabar three times. Second, copying is reciprocal:[26] @BugungiGap and @Yaxshi_Yomon copied each other (3 and 2 times), as did @patrul_uz and @qoraxabar (5 and 3 times). This is the signature of mutual monitoring among competitors, not of a hierarchy. Third, official sources sit upstream of professional media: the presidential press service published the Li Qiang meeting 45 minutes before Kun.uz. In the five three-post families, timing alone cannot show whether the third channel copied the root or the second post, and with families of at most three posts chain and star structures cannot be statistically distinguished. H3 is not supported: the observed structure is a set of reciprocal peer clusters fed by official sources, not a single-hub star.

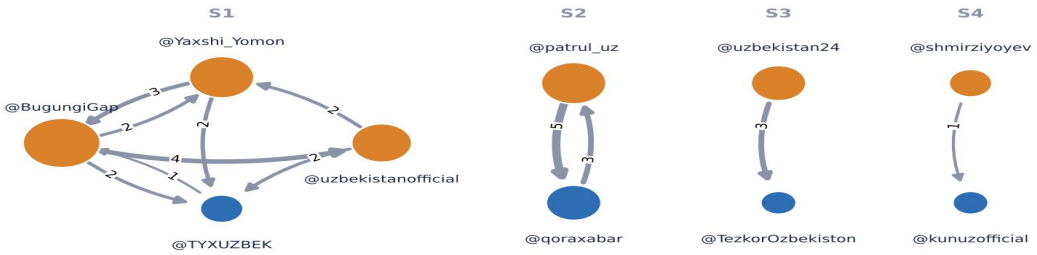


Figure 5. Observed republication network (root → copier), by publication day[27]

The forward metadata is equally telling. Of the 264 posts read, only five carried a "Forwarded from" or "via" label, and these were all advertisements, bots or the channel's own earlier posts. None was a forward from another news channel. All republication of news therefore took place by copy-and-paste without visible attribution, consistent with the 3.4% attribution rate in France. Finally, roots reached a larger share of their channel's audience than copies in 64% of families, with a mean log difference of 0.13 (about 14%). The difference is not significant at the 5% level (p = 0.06), so H4 receives only weak support.

5. Discussion

Distribution without production. The central finding is the gap between audience and source status. Telegram has let many channels build audiences of several hundred thousand to two million subscribers, but citations, our proxy for originating content, are concentrated in a narrower group led by a professional newsroom (Kun.uz) and a set of fast aggregators. The bottom half of large

[26]Reciprocal copying. A dyadic pattern in which each of two channels sometimes publishes first and is sometimes copied by the other. In network terms, it produces edges in both directions between the same pair of nodes.

[27] Source: author's coding of Telegram public web previews. Edge labels give the number of families; orange nodes published first more often than they copied.

channels holds a fifth of subscribers but less than a tenth of citations. This pattern is consistent with the idea of an attention economy:[28] value accrues to channels that capture attention, not necessarily to those that report. Because none of the observed republications carried a native forward, originators receive no visible credit when their stories are reused.

Speed and verification. The median first copy appeared within about 20 minutes, and some channel pairs mirrored each other within minutes. Such delays leave little time for independent verification and fit the logic of information cascades (Bikhchandani, Hirshleifer and Welch, 1992). The reciprocal structure is a mixed blessing. It avoids the single point of failure of a pure star, but tightly coupled peers can amplify each other's errors and make it hard for readers to identify the original source.

Implications. First, for platform and industry norms, the gap between copy-and-paste and native forwarding suggests that attribution standards, for example a media association code requiring a source link, would return audience value to originators at little cost. Second, for newsrooms, a modest and short-lived first-mover premium (H4) suggests investing in formats that are harder to copy, such as exclusive video, data and analysis. Third, for regulators and **researchers**, the TGStat citation index offers a cheap, transparent indicator for monitoring source concentration over time, which could complement audience-based concentration measures in media policy.

Scaling up. The story-tracing component is necessarily small. Figure 6 outlines the pipeline we propose for a full-scale replication. It would collect all posts from 100–150 channels over two to three months with the open-source Telethon library, encode them with a multilingual sentence encoder[29] that handles Latin, Cyrillic and Russian text,[30] cluster near-duplicates into families, and attribute each copy to its nearest earlier predecessor. Attributing every copy to the earliest post imposes a star structure by construction, so the nearest-predecessor rule is needed to estimate topology rather than assume it.

[28]**Attention economy.** The idea, going back to Herbert A. Simon (1971), Nobel laureate in economics (1978), that "a wealth of information creates a poverty of attention": when information is abundant, audience attention becomes the scarce resource that outlets compete for.

[29]**Sentence embedding.** A fixed-length numerical vector produced by a neural language model to represent a text's meaning; paraphrases, translations and transliterations of the same story receive nearby vectors. Multilingual encoders such as LaBSE (Feng et al., 2022) cover Uzbek and Russian.

[30]**Uzbek scripts.** Uzbekistan adopted a Latin-based alphabet in 1993, but Cyrillic remains in wide use, and many channels publish in Latin Uzbek, Cyrillic Uzbek and Russian. Identical stories therefore often differ in surface form.

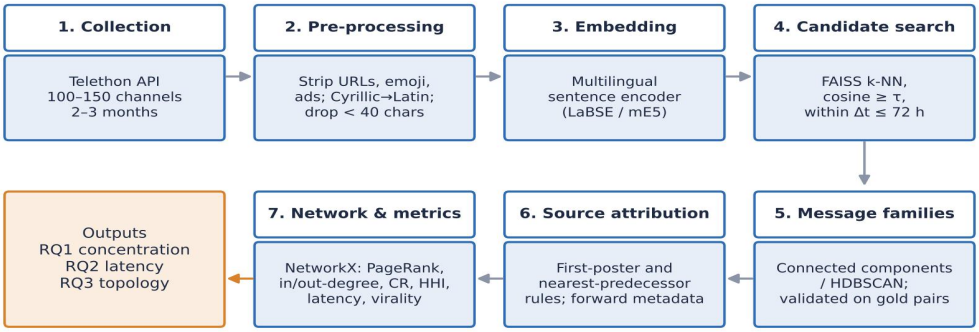


Figure 6. Proposed embedding-based pipeline for large-scale replication
τ = similarity threshold; Δt = maximum time window within a family.

Limitations. The citation index is a proprietary proxy that may include paid promotional mentions. Dataset A covers only the 100 largest channels, so the full tail is even less cited. The traced sample is small, drawn from cached pages covering four days, and matched manually. The "first observed" publisher may not be the true origin, since stories can originate on websites or in channels outside the sample. Times are shown to the minute and may reflect scheduled posting. The results should therefore be read as exploratory evidence that motivates, and is consistent with, a larger study.

6. Conclusion

This paper set out to measure how centralised information flow is in Uzbekistan's Telegram news ecosystem. Three findings emerge. Concentration: source status is markedly more concentrated than audience (Gini 0.61 vs. 0.46), but no small group dominates. The top ten channels receive 39% of citations, led by Kun.uz, and audience size is only a partial guide to source status.

Latency: stories are typically first republished within about 20 minutes, and some channel pairs mirror each other within minutes.

Structure: republication takes the form of reciprocal copying among peer channels fed by official sources, entirely without native attribution, rather than a single-hub star. For the economics of news, these findings indicate a market in which distribution has multiplied far faster than original production, and in which the returns to being first are weak and largely invisible. Attribution norms and large-scale monitoring with the embedding-based pipeline proposed here are the natural next steps.

Declarations

Data availability. Dataset A (TGStat ranking, 100 channels) and Dataset B (25 coded story families) are available from the author on request. **Ethics.** Only public posts by media channels were analysed at the channel level; no data on individual users were collected. **Conflict of interest.** The author declares none.

Appendix A. Traced story families

ID	Day	Story	Sequence (channel and time shown)	Latency (min)

ID	Day	Story	Sequence (channel and time shown)	Latency (min)
F01	S1	School security regime: checkpoints and electronic entry logs	@BugungiGap 09:59 → @Yaxshi_Yomon 10:40 → @TYXUZBEK 12:07	41, 128
F02	S1	Weather: sharp cooling then return of heat in August	@Yaxshi_Yomon 11:07 → @BugungiGap 12:59	112
F03	S1	Two animals introduced as Presidential Administration 'staff'	@uzbekistanofficial 07:59 → @Yaxshi_Yomon 12:30 → @TYXUZBEK 12:54	271, 295
F04	S1	Child's eyesight falls to -8 after phone use	@Yaxshi_Yomon 14:06 → @BugungiGap 14:15 → @TYXUZBEK 15:23	9, 77
F05	S1	Party proposes licence revocation for speeding >80 km/h	@BugungiGap 14:44 → @uzbekistanofficial 14:52 → @Yaxshi_Yomon 14:56	8, 12
F06	S1	Drifting may lead to vehicle confiscation	@Yaxshi_Yomon 15:51 → @BugungiGap 16:00	9
F07	S1	Heating subsidy proposal for flats over 60 sq m	@BugungiGap 09:51 → @uzbekistanofficial 13:16	205
F08	S1	Dealer discounts on Cobalt and Damas	@BugungiGap 09:25 → @uzbekistanofficial 14:25	300
F09	S1	Drunk driver flees traffic inspector in Chirchiq	@TYXUZBEK 13:14 → @BugungiGap 13:20	6
F10	S1	Energy minister: cheaper tariff for areas without gas	@BugungiGap 15:37 → @TYXUZBEK 16:18	41
F11	S1	Traffic inspector killed by drunk driver in Fergana	@uzbekistanofficial 07:26 → @TYXUZBEK 17:10 → @Yaxshi_Yomon 17:48	584, 622
F12	S1	Mudflow and flood warning for 4-7 August	@Yaxshi_Yomon 16:25 → @TYXUZBEK 18:12	107
F13	S1	Viral wedding video: groom's party at bride's home	@BugungiGap 10:53 → @uzbekistanofficial 13:12	139
F14	S2	Brawl at Yunusobod Universam bus stop	@patrul_uz 12:55 → @qoraxabar 13:03	8
F15	S2	New head of Andijan regional police appointed	@patrul_uz 13:02 → @qoraxabar 13:13	11
F16	S2	Narcotics worth over $600000 seized in Tashkent	@qoraxabar 13:33 → @patrul_uz 13:35	2

ID	Day	Story	Sequence (channel and time shown)	Latency (min)
F17	S2	Six-year General Medicine bachelor's programme proposed	@patrul_uz 13:53 → @qoraxabar 14:28	35
F18	S2	Filters remove odour from Salar wastewater plant	@qoraxabar 14:50 → @patrul_uz 15:37	47
F19	S2	Document to standardise salaries across ministries	@patrul_uz 15:39 → @qoraxabar 15:58	19
F20	S2	School director sentenced for selling lesson hours	@patrul_uz 16:01 → @qoraxabar 16:13	12
F21	S2	Satirical post on Ai-95 petrol price	@qoraxabar 16:41 → @patrul_uz 16:47	6
F22	S3	Heavy snow in Qorakoy area of Kyrgyzstan	@uzbekistan24 12:49 → @TezkorOzbekiston 12:55	6
F23	S3	Rules for Green Card lottery applications	@uzbekistan24 14:00 → @TezkorOzbekiston 14:13	13
F24	S3	New 780-m bridge to Samarkand airport	@uzbekistan24 14:35 → @TezkorOzbekiston 14:55	20
F25	S4	President receives Chinese delegation led by Li Qiang	@shmirziyoyev 15:45 → @kunuzofficial 16:30	45

Source: author's coding of Telegram public web previews retrieved on 18 October 2026. Latency is measured from the root post.

References

1. Baumgartner, J., Zannettou, S., Squire, M. and Blackburn, J. (2020) 'The Pushshift Telegram dataset', Proceedings of the International AAAI Conference on Web and Social Media, 14, pp. 840–847.
2. Bikhchandani, S., Hirshleifer, D. and Welch, I. (1992) 'A theory of fads, fashion, custom, and cultural change as informational cascades', Journal of Political Economy, 100(5), pp. 992–1026.
3. Boczkowski, P.J. (2010) News at Work: Imitation in an Age of Information Abundance. Chicago: University of Chicago Press.
4. CABAR.asia (2020) IWPR research: social networks are an important source of news in Uzbekistan. Available at: https://cabar.asia/?p=30589 (Accessed: 18 October 2026).
5. Cagé, J., Hervé, N. and Viaud, M.-L. (2020) 'The production of information in an online world', The Review of Economic Studies, 87(5), pp. 2126–2164. doi: 10.1093/restud/rdz061.
6. Chen, X., Zeynali, A., Camargo, C., Flöck, F., Gaffney, D., Grabowicz, P., Hale, S.A., Jurgens, D. and Samory, M. (2022) 'SemEval-2022 Task 8: Multilingual news article similarity', in

Proceedings of the 16th International Workshop on Semantic Evaluation. Seattle: ACL, pp. 1094–1106.
7. Davies, N. (2008) Flat Earth News. London: Chatto & Windus.
8. Donaev, M. (2024) How the news goes viral in Uzbekistan. Global Voices, 14 October. Available at: https://globalvoices.org/2024/10/14/how-the-news-goes-viral-in-uzbekistan/ (Accessed: 18 October 2026).
9. Efron, B. (1979) 'Bootstrap methods: another look at the jackknife', The Annals of Statistics, 7(1), pp. 1–26.
10. Feng, F., Yang, Y., Cer, D., Arivazhagan, N. and Wang, W. (2022) 'Language-agnostic BERT sentence embedding', in Proceedings of the 60th Annual Meeting of the ACL. Dublin: ACL, pp. 878–891.
11. Gentzkow, M. and Shapiro, J.M. (2008) 'Competition and truth in the market for news', Journal of Economic Perspectives, 22(2), pp. 133–154.
12. Gini, C. (1912) Variabilità e mutabilità. Bologna: Tipografia di Paolo Cuppini.
13. Goel, S., Anderson, A., Hofman, J. and Watts, D.J. (2016) 'The structural virality of online diffusion', Management Science, 62(1), pp. 180–196.
14. Gomez-Rodriguez, M., Leskovec, J. and Krause, A. (2012) 'Inferring networks of diffusion and influence', ACM Transactions on Knowledge Discovery from Data, 5(4), Article 21.
15. Hamilton, J.T. (2004) All the News That's Fit to Sell. Princeton: Princeton University Press.
16. Harder, R.A., Sevenans, J. and Van Aelst, P. (2017) 'Intermedia agenda setting in the social media age', The International Journal of Press/Politics, 22(3), pp. 275–293. doi: 10.1177/1940161217704969.
17. Hirschman, A.O. (1964) 'The paternity of an index', The American Economic Review, 54(5), p. 761.
18. Kepios (2026) Digital 2026: Uzbekistan. DataReportal. Available at: https://datareportal.com/reports/digital-2026-uzbekistan (Accessed: 18 October 2026).
19. Kun.uz (2023) Uzbekistan ranks second in the world in terms of number of Telegram channels. Available at: https://kun.uz/en/50659966 (Accessed: 18 October 2026).
20. Leskovec, J., Backstrom, L. and Kleinberg, J. (2009) 'Meme-tracking and the dynamics of the news cycle', in Proceedings of the 15th ACM SIGKDD Conference. Paris: ACM, pp. 497–506.
21. Lewis, J., Williams, A. and Franklin, B. (2008) 'A compromised fourth estate? UK news journalism, public relations and news sources', Journalism Studies, 9(1), pp. 1–20.
22. Lorenz, M.O. (1905) 'Methods of measuring the concentration of wealth', Publications of the American Statistical Association, 9(70), pp. 209–219.
23. McCombs, M.E. and Shaw, D.L. (1972) 'The agenda-setting function of mass media', Public Opinion Quarterly, 36(2), pp. 176–187.
24. Nicholls, T. (2019) 'Detecting textual reuse in news stories, at scale', International Journal of Communication, 13. Available at: https://ijoc.org/index.php/ijoc/article/view/9904 (Accessed: 18 October 2026).
25. Reimers, N. and Gurevych, I. (2020) 'Making monolingual sentence embeddings multilingual using knowledge distillation', in Proceedings of EMNLP 2020. Online: ACL, pp. 4512–4525.
26. Samuelson, P.A. (1954) 'The pure theory of public expenditure', The Review of Economics and Statistics, 36(4), pp. 387–389.
27. Simon, H.A. (1971) 'Designing organizations for an information-rich world', in Greenberger, M. (ed.) Computers, Communications, and the Public Interest. Baltimore: Johns Hopkins Press, pp. 37–72.

28. TGStat (2026) Telegram kanallar reytingi: O‘zbekiston — Yangiliklar va OAV. Available at: https://uz.tgstat.com/ratings/channels/news (Accessed: 18 October 2026).
29. Urman, A. and Katz, S. (2022) ‘What they do in the shadows: examining the far-right networks on Telegram’, Information, Communication & Society, 25(7), pp. 904–923.
30. Vliegenthart, R. and Walgrave, S. (2008) ‘The contingency of intermedia agenda setting: a longitudinal study in Belgium’, Journalism & Mass Communication Quarterly, 85(4), pp. 860–877.
31. Vosoughi, S., Roy, D. and Aral, S. (2018) ‘The spread of true and false news online’, Science, 359(6380), pp. 1146–1151.
32. Wijermars, M. and Lokot, T. (2022) ‘Is Telegram a “harbinger of freedom”? The performance, practices, and perception of platforms as political actors in authoritarian states’, Post-Soviet Affairs, 38(1–2), pp. 125–145.
33. Wilcoxon, F. (1945) ‘Individual comparisons by ranking methods’, Biometrics Bulletin, 1(6), pp. 80–83.