nature communications

Article

https://doi.org/10.1038/s41467-025-64785-1

Temporal fingerprints for identity matching across fully encrypted domains

Received: 25 February 2025

Accepted: 25 September 2025

Published online: 27 October 2025

Check for updates

Shahar Somin[1,2] ✉, Keeley Erhardt[1,3], Tom Cohen[1,3], Jeremy Kepner[1] & Alex 'Sandy' Pentland[1]

In the digital age, coordinated inauthentic behavior threatens societal stability, markets, and security. Advances in generative AI amplify these threats, enabling effortless content creation, amplifying actors' influence. Detection is hindered by cross-domain activity, where pseudonymous profiles operate across encrypted platforms, and by privacy constraints limiting content analysis. In this study, we propose a robust and scalable cross-domain identity matching framework, based on bursty dynamics, independent of content or interaction data. It outperforms state-of-the-art temporal and structural approaches, remains resilient to incomplete data, and correctly identifies 35% of profiles after 52 weeks. It scales effectively, attaining AUC 0.78 when matching identities across 500 marketplaces with over 250k daily traders. By framing identity matching within the "network of networks" perspective, we demonstrate how coordinated behavior propagates across domains. This dual methodological and theoretical contribution paves the way for innovative strategies to combat digital threats in an increasingly complex and adversarial landscape.

In an era characterized by digital ubiquity, the nature of human interaction has undergone a profound transformation. Historically, our interactions, whether social or commercial, relied on physical encounters, bound to a singular persona—our physical identity. However, with the emergence of online digital platforms, the vast majority of our activities occur online, facilitating the use of different profiles for diverse objectives and platforms. The rapid advancements in generative artificial intelligence (AI) and large language models (LLMs) have further impacted our digital environments, making content creation effortless and instantaneous. In this digitally dispersed and under-regulated landscape, malicious or deceptive use of multiple online identities has become a growing concern. Indeed, governments, corporations, and nations alike face serious threats as entities exploit anonymity and platform diversity to manipulate narratives, influence public opinion, and disrupt markets, posing risks to societal stability, economic integrity[1], and national security[2–4]. Detecting and mitigating such threats across platforms presents considerable challenges, as malicious activity is often concealed and fragmented across pseudonymous identities. Addressing these challenges requires robust methodologies for linking profiles associated with the same real-world entities across domains, referred to as the identity matching problem (Network alignment or user identity linkage are terminologies often used for this problem as well).

Extensive research efforts have demonstrated that content and personal attributes, such as age, gender or usernames are effective for matching profiles across domains. These approaches include both supervised[5–9] and unsupervised[10–13] models. However, content-based models face significant challenges nowadays. First, the widespread use of generative AI enables creating numerous variations of the same narrative effortlessly, making it increasingly difficult to detect content originating from the same individual. Additionally, the increasing shift toward encrypted platforms often restricts access to user-generated content and private attributes, especially across different platforms, further limiting the effectiveness of these models.

Furthermore, even in anonymized settings, notable cross-domain identity matching capabilities have been achieved by exploiting the network of interactions[14–23] or individual metadata, such as profile trajectories[24]. However, such structure-based models face several

[1]MIT, Cambridge, MA, USA. [2]Bar-Ilan University, Ramat-Gan, Israel. [3]These authors contributed equally: Keeley Erhardt, Tom Cohen.
✉e-mail: shaharso@mit.edu

limitations. Different platforms often exhibit distinct types of connections, leading to unique structural patterns that may hinder the effectiveness of these methods. Additionally, the sheer volume of structural information, especially when considering k-hop neighbors and the dynamic nature of connections, makes comprehensive data collection a difficult task. Even when structural data is completely available, running such models at scale is computationally expensive and even infeasible at times.

Building on the concept of "network of networks"[25–27], we propose that distinct domains are implicitly connected through profiles controlled by the same entity. These hidden connections enable external events to propagate like shock waves, influencing actions across otherwise disconnected platforms. The temporal traces left by these actions, such as the timing of phone calls, digital transactions, or social media activity, offer measurable indicators of the underlying connections, bridging the gap between implicit links and explicit observable data. Bursty patterns of human behavior[28], represented by the gap between consecutive activities of the individual (inter-event times), offer a distinct perspective of human dynamics, emphasizing the frequency of activity, rather than precise timing of individual events. These distributions were previously encountered in many and diverse types of activities, including individual mobility patterns, e-mail communications, instant messaging, web browsing, and mobile phone calls[29–34].

In this study, we demonstrate that beyond conforming to a heavy-tailed distribution, bursty patterns are uniquely personal and can effectively characterize an individual, even upon acting across different platforms. Interestingly, these patterns provide a distinctive signature, better than the network of interactions and the actual timing of activity, enabling the detection of multiple profiles corresponding to the same individual, across different encrypted domains. We demonstrate the bursty model's performance on two use-cases: across different financial platforms and across different social platforms. First, we show that our model outperforms various state-of-the-art temporal and structural models, presenting an average area under the receiver operating characteristic (ROC) curve (AUC) of 0.86, across two examined financial marketplaces. We further demonstrate its high stability over time, enabling the correct identification of 35% of profiles even after an entire year, by examining at most 10 candidates for each profile. This suggests that the temporal fingerprints last not only across domains but also for long periods of time. We additionally evaluate the model's scalability, applying it across 500 different market places, encompassing the activity of over 250k daily traders. The identity matching problem across these domains presents a notably low baseline, as merely 3 out of 1000 randomly selected profile pairs actually correspond to the same individual. Nevertheless, our methodology achieves an average AUC of 0.78 and precision of 96% for the top-100 predictions. Finally, we show that the model is generalizable to other types of domains, establishing an AUC of 0.63 for matching identities across Twitter, Telegram and Instagram data, and achieving AUCs of 0.77 and 0.89 for matching identities across different subreddits and across different Telegram channels, respectively.

As the digital world enables the effortless creation of different content based on the same narrative content-based identity matching models become increasingly hindered. In parallel, the growing adoption of privacy-preserving measures further restrict access to identifying content. Accurately matching identities across platforms under these constraints becomes more challenging, but is inherently essential for uncovering patterns that extend across fragmented digital environments. By leveraging bursty individual patterns, our model provides a robust solution for linking identities within and across encrypted and content-restricted domains. Additionally, compared to other state-of-the-art temporal and structural methods with significantly higher computational complexity, our approach offers a vital advantage in enabling fast and accurate detection at scale. By enabling accurate identity matching across domains, even when content and network structure are absent, the proposed methodology provides a critical building block for detecting coordinated behavior across fragmented systems. Coupled with the suggested broader perspective on the mechanisms driving cross-domain coordination, this work lays a robust foundation for advancing threat detection methodologies in a rapidly evolving digital landscape, where traditional methods often fall short.

Results

Preliminaries

In this study, we aim at identifying individuals across different encrypted domains. Specifically, we aim at learning an identity matching function:

Definition 1. *Given $D_1, \ldots, D_n$ different domains, the goal of the cross-domain identity matching problem is learning a function:*

$$p : \bigcup_{i,j\in[n]} D_i \times D_j \to [0,1]$$

such that $p(u_{d_1}, v_{d_2})$ represents the probability that the profiles $u_{d_1} \in D_1$ and $v_{d_2} \in D_2$ are associated with the same real-world individual ($u_{d_1} = v_{d_2}$).

The vanilla inter-event bursty model. We propose exploiting individual temporal data for linking profiles back to the same individual across different domains. Specifically, we analyze individual bursty patterns, manifested as the time difference between any two consecutive activities of each profile. Formally:

Definition 2. *Given a time period $[\tau, \tau + \Delta\tau]$ and a profile u_d in domain D, we denote the sequence of their activity times $A_\tau^{u_d} \subset [\tau, \tau + \Delta\tau]$ by:*

$$A_\tau^{u_d} = (t_0^{u_d}, \ldots, t_m^{u_d}) \tag{1}$$

An inter-event time period is defined as the time difference between two consecutive activities of u_d:

$$\Delta t_i^{u_d} = t_i^{u_d} - t_{i-1}^{u_d} \tag{2}$$

The inter-event time sequence is defined by:

$$S_\tau^{u_d} = (\Delta t_1^{u_d}, \ldots \Delta t_m^{u_d}) \tag{3}$$

The cumulative distribution function of the inter-event sequence is defined as:

$$Q_\tau^{u_d}(\Delta t) = \frac{|\delta \in S_\tau^{u_d} : \delta \le \Delta t|}{m} \tag{4}$$

Our bursty identity matching function p^{ks} is based on the similarity between the established inter-event time distributions of any two profiles, estimated by the Kolmogorov–Smirnov (KS) statistic:

Definition 3. *Let $u_{d_1} \in D_1^\tau$ and $v_{d_2} \in D_2^\tau$, and their corresponding inter-event time distributions $Q_\tau^{u_{d_1}}$ and $Q_\tau^{v_{d_2}}$. The KS-statistic is defined as the maximal difference between their distributions:*

$$KS_\tau(u_{d_1}, v_{d_2}) = \sup_{\Delta t} |Q_\tau^{u_{d_1}}(\Delta t) - Q_\tau^{v_{d_2}}(\Delta t)| \tag{5}$$

We define the corresponding identity matching function p_τ^{ks} as:

$$p_\tau^{ks}(u_{d_1}, v_{d_2}) = 1 - KS_\tau(u_{d_1}, v_{d_2}) \tag{6}$$

https://doi.org/10.1038/s41467-025-64785-1

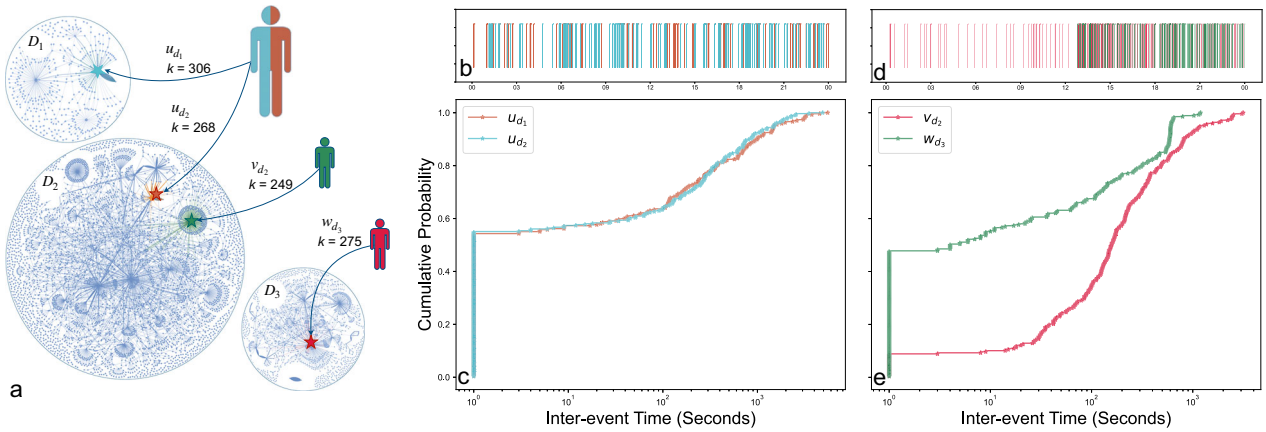

Fig. 1 | Synchronization of burstiness patterns. a illustrates the daily networks of three financial trading markets D_τ^1, D_τ^2 and D_τ^3, each corresponding to the trading of a different crypto-token. Profiles u_{d_1} (degree $k = 306$) and u_{d_2} (degree $k = 268$) correspond to the same individual (illustrated by orange and cyan markers). Profiles v_{d_2} (degree $k = 249$) and w_{d_3} (degree $k = 275$) pertain to different individuals (illustrated by red and green markers). **b** presents the activity times of u_{d_1} and u_{d_2}, reaching an activity overlap of 37%. **d** presents the activity times of v_{d_2} and w_{d_3}, reaching an activity overlap of 42%. **c** depicts the cumulative inter-event times distributions of u_{d_1} and u_{d_2}, exemplifying similar distributions ($KS_\tau(u_{d_1}, u_{d_2}) = 0.031$ with a p value of 0.99). **e** depicts the cumulative inter-event times distributions of v_{d_2} and w_{d_3}, exemplifying significantly different distributions ($KS_\tau(v_{d_2}, w_{d_3}) = 0.47$ with a p-value of 5e−27).

We postulate that different profiles pertaining to the same individual are bound to exhibit synchronization in their activity dynamics, despite acting on different domains. The motivation to this hypothesis is presented in Fig. 1, observing two pairs of profiles:

1. A positive pair: u_{d_1} and u_{d_2}, correspond to the same individual u in financial trading markets D_τ^1 and D_τ^2 respectively, illustrated illustrated in Fig. 1a (orange and cyan markers).
2. A negative pair: v_{d_2} and w_{d_3}, corresponding to different individuals in two different financial trading markets $v \in D_\tau^2$, $w \in D_\tau^3$, illustrated by red and green markers in Fig. 1a.

Both pairs present similar degrees (illustrated in Fig. 1a) and resembling overlap in activity times: 37% of the time for the positive pair and 42% for the negative pair, illustrated in Fig. 1b, d, respectively. Nevertheless, the negative pair presents significantly different inter-event time distributions, establishing a KS distance of $KS_\tau(v_{d_2}, w_{d_3}) = 0.47$ with a p value of $5e^{-27}$ (Fig. 1e), while the positive pair presents a high similarity of the inter-event distributions, with $KS_\tau(u_{d_1}, u_{d_2}) = 0.031$ and a p-value of 0.99 (illustrated in Fig. 1c). This illustrates that burstiness patterns are able to characterize the profiles pertaining to the same individual better than basic temporal and structural patterns.

Experiments on financial markets

Identifying collusive trading practices, fraudulent accounts, and coordinated market manipulations, often concealed under multiple pseudonymous profiles to evade detection, is essential for ensuring the integrity of financial systems and maintaining market stability. We examine $2k$ traders transacting on two different financial trading markets, on top of the Ethereum blockchain[35–37]. In this setting, we refer to a financial trading market D_τ^i as encompassing all of the trading activity related to the respective crypto-token c_i and time period $[\tau, \tau + \Delta\tau]$ where $\Delta\tau$ stands for single day length:

$$D_\tau^i = \{u : u \text{ bought or sold } c_i \text{ in } [\tau, \tau + \Delta\tau]\} \tag{7}$$

We wish to evaluate the performance of the vanilla inter-event bursty model and compare it to baseline models exploiting structural[14,15,19,20] and temporal[38–41] node characteristics (consider section "Methods" for a formal definition). Figure 2 presents the performance analysis of these models. Specifically, we consider 14 days of activity over two examined financial trading markets D_τ^1, D_τ^2 and evaluate the performance on each day separately. Figure 2a presents the comparison of the AUC (ROC curve), and Fig. 2b depicts comparison of the precision established for each threshold of top-ranked profile pairs (u_{d_1}, v_{d_2}). Both metrics indicate that structure-based models demonstrates limited efficacy in linking profiles across different domains, highlighting the greater significance of temporality in this context (consider section "Discussion" for a thorough discussion). The vanilla inter-event bursty model outperforms all temporal baselines, suggesting that while the bursty dynamics are closely related to an individual's actual activity times, the latter is less effective in capturing the nuances required for an accurate individual fingerprint.

Stability and robustness

We further wish to evaluate the ability to match profiles across domains even after long periods of time. In particular, given a user u with a profile $u_{d_1} \in D_1^{t_0}$, active during $[t_0, t_0 + \Delta\tau]$ in domain D_1, we define the similarity of u_{d_1} to other profiles in the second domain, after a time delay τ as:

$$KS_{t_0,\tau}(u_{d_1}, v_{d_2}) = \sup_{\Delta t} |Q_{t_0}^{u_{d_1}}(\Delta t) - Q_{t_0+\tau}^{v_{d_2}}(\Delta t)| \tag{8}$$

The corresponding identity matching function is:

$$p_{t_0,\tau}^{ks}(u_{d_1}, v_{d_2}) = 1 - KS_{t_0,\tau}(u_{d_1}, v_{d_2}) \tag{9}$$

We define the identification probability of u at time τ as the probability that $p_{t_0,\tau}^{ks}(u_{d_1}, u_{d_2})$ is within the top-k ranked matches for u_{d_1}. Figure 3a, b present the identification probability as a function of the examined time delay τ, for $k = 5$ and $k = 10$ correspondingly. While the identification probability decreases with the length of delay, the performance remains rather high with 25% (35%) of users correctly identified within 5 (10) matches, even after a year long, outperforming baseline temporal models.

Next we examine the robustness of our suggested model to data omission. Indeed, incomplete data is a common challenge in today's era of massive data generation, making it crucial to understand its implications. We randomly omit Δ of each profile's activity times within the different domains and analyze the omission effect. Figure 4a presents the cumulative inter-event time distributions of two examined profiles, which pertain to the same individual, after randomly omitting

https://doi.org/10.1038/s41467-025-64785-1

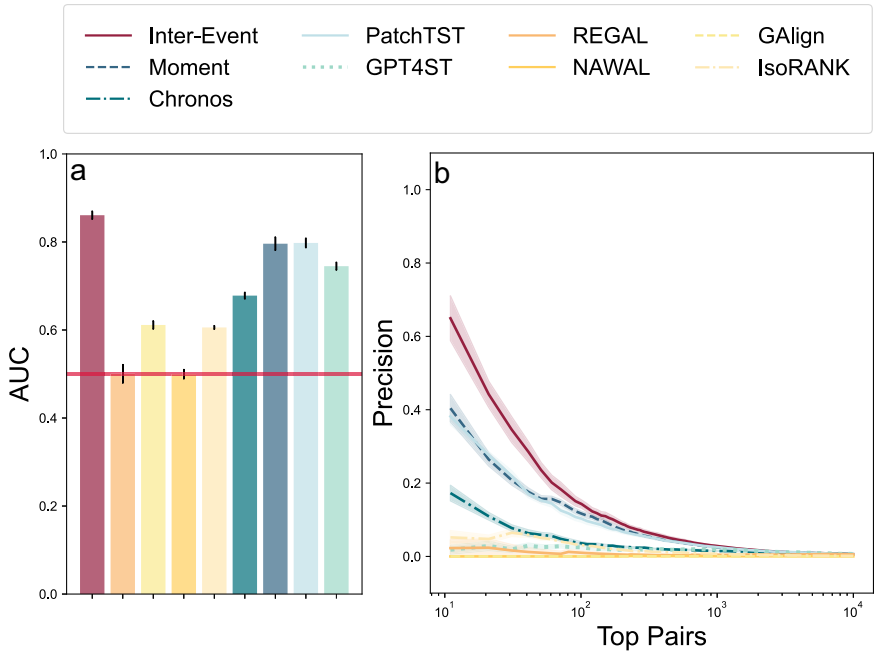

Fig. 2 | Performance evaluation for the cross-domain identity matching problem. Comparing the inter-event bursty model (p^{ks}) and temporal (blue shaded) and structural (yellow shaded) baselines. **a** presents the averaged AUC over 14 daily tests, with error bars standing for standard error. **b** presents the average precision as a function of the examined number of pair candidates (with ±1 standard error in light background, correspondingly). The inter-event bursty model presents higher performance than baseline methods across all examined metrics.

$\Delta \in [50\%, 75\%, 90\%]$ of their activity, compared to their original distributions. Notably, the pre- and post-data omission distributions change significantly, and the similarity between the two profiles decreases with Δ. Figure 4b depicts the change to the average KS distance as a function of the original KS, presenting a non-monotonous effect of data omission on profile similarity. Encouragingly, this indicates that both originally highly similar and highly distinct profiles are more robust to data omission, compared to profile pairs that originally had medium-level similarity, and are more prone to be affected by incomplete data. We further wish to estimate the data omission effect on the performance of the identity matching model. Figure 4c, d depict correspondingly the AUC and precision for each of the different levels of Δ, indicating the decrease caused by incomplete data. Despite this expected decrease in performance, the vanilla inter-event bursty model, even under 90% data omission threshold, outperforms state-of-the-art temporal models, applied with only 50% omission threshold (consider Supplementary Fig. 3). This result underscores the model's relative robustness and its effectiveness in capturing individual fingerprints, even under severe data limitations.

Scalability: Bursty-GNN for temporal similarity networks

To further enhance the vanilla version of inter-event bursty model, we propose employing a temporal graph neural network (TGNN) on top of cross-domain similarity networks $G^{\tau}_{ks} = (V^{\tau}, E^{\tau})$ where edges link profiles across different domains, and edge weight is reflected by the KS distance between the inter-event time distributions of any pair of profiles. We employ a supervised learning approach to train the TGNN, using edge labels inferred from the KS statistic metric:

1. Positive edges: two profiles $u_{d_1} \in D^{\tau}_1$ and $v_{d_2} \in D^{\tau}_2$ are linked by a positive edge if $KS_{\tau}(u_{d_1}, v_{d_2}) \le th^p$, where th^p is a predefined positive threshold.
2. Negative edges: two profiles $u_{d_1} \in D^{\tau}_1$ and $v_{d_2} \in D^{\tau}_2$ are linked by a negative edge if $KS_{\tau}(u_{d_1}, v_{d_2}) \ge th^n$, where th^n is a predefined negative threshold.

We employed $th^p = 0.001$ and $th^n = 0.98$ as the predefined positive and negative thresholds, respectively. The proposed TGNN setting utilizes labels inferred from the KS statistic and does not rely on actual identity labels. As such, it is applicable to the unsupervised setting we are examining. The Bursty-TGNN learns a latent embedding for all profiles, which is utilized subsequently for a cross-domain edge detection task. Figure 5a illustrates the two-layer Bursty-TGNN employed on the daily similarity networks. An elaborated overview of the TGNN architecture can be found in the Supplementary Materials.

In order to evaluate the Bursty-TGNN, we consider an identity matching problem of higher complexity, where we do not restrict the experiment to the two-domains use-case. This setting, alongside verifying the performance of the Bursty-TGNN, will assist in examining its scalability by identifying profiles across over 500 financial trading markets, while considering over 250k daily users. The performance was evaluated for the vanilla bursty model, the Bursty-TGNN model comparing them with temporal baseline models only. The structural baseline models did not scale effectively, preventing us from evaluating their performance in this challenge.

The identity matching problem across these domains presents a notably low baseline, as merely 3 out of 1000 randomly selected profile pairs actually correspond to the same individual (dashed horizontal red line in Fig. 5c). Figure 5b, c depict, respectively, the average AUC and precision for the multi-domain setting. Notably, up to the top-200 pairs the vanilla inter-event bursty model outperforms the temporal baselines, reaching almost error-less precision on average. Furthermore, the Bursty-TGNN extension (dashed purple curve, Fig. 5c) presents a performance boost when run on the top-1000 inter-event similarity edges (dashed vertical gray line in Fig. 5c marks the top-1000 threshold), underscoring the evolving role of each profile within the overall network and the hidden potential in this dynamic view for solving the identity matching task.

https://doi.org/10.1038/s41467-025-64785-1

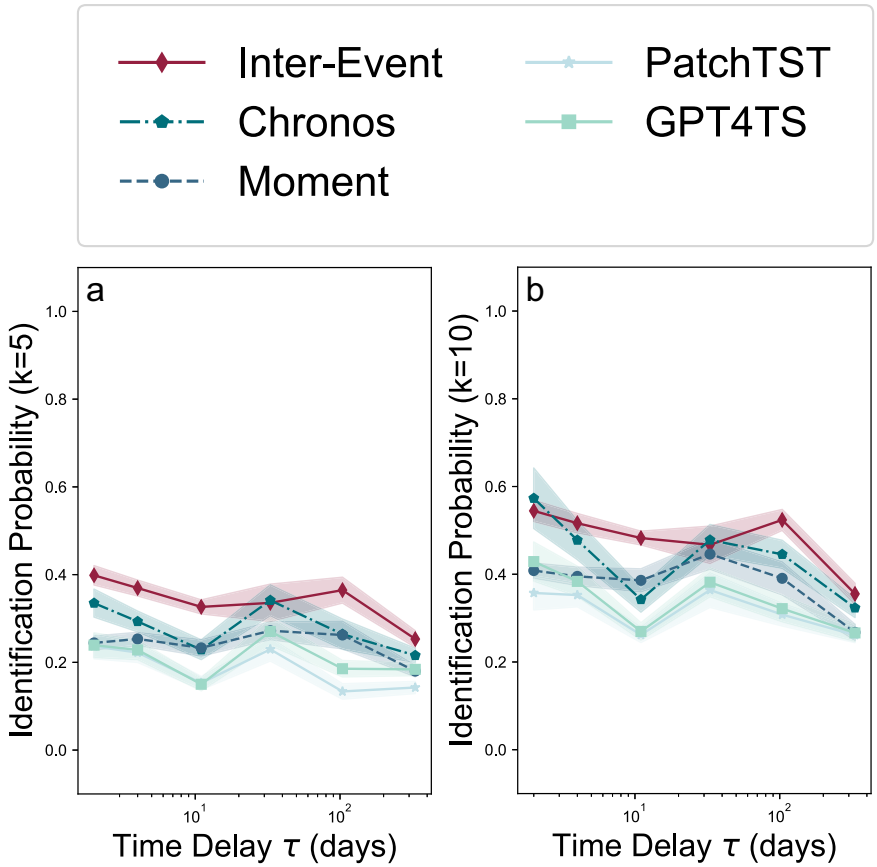

Fig. 3 | Temporal fingerprints stability. The identification probability of a user depending on the time delay from their initial observation, for $k = 5$ and $k = 10$, in (**a**, **b**) correspondingly. Despite the evident decrease of the identification probability as the delay between observations increases, the vanilla bursty model (dark red curve) is able to correctly identify 25% of the users within 5 ranks, and 35% within 10 ranks, after a delay of an entire year, outperforming the baseline temporal models (green-shaded curves). Shaded background represents ±1 standard error.

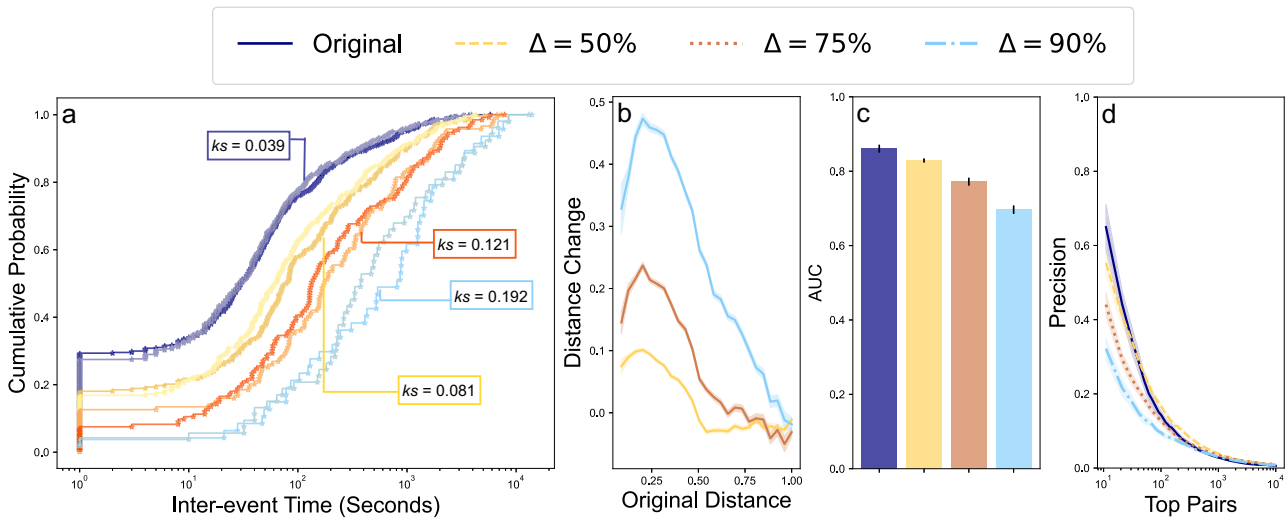

Fig. 4 | Robustness to incomplete data. a depicts the cumulative inter-event time distributions of two examined profiles, which pertain to the same individual, after randomly omitting $\Delta \in [50\%, 75\%, 90\%]$ of their activity, compared to their original distributions. Notably, the similarity deteriorates with increase in Δ. **b** presents the non-monotonous effect of data omission on the average KS distance change as a function of the original KS distance. **c**, **d** present the AUC and the precision of p^{ks} upon different noise omission thresholds, presenting a slight decrease in performance. Error bars (**c**) and light shaded background (**b**, **d**) represent standard error.

Experiments on social media

Identifying coordinated inauthentic behavior across social media domains is essential, since malicious entities frequently employ bots, troll farms, or fake accounts to avoid detection and reinforce targeted narratives. These tactics are often aimed at shaping public sentiment, influencing elections, and steering political landscapes. Platforms like Twitter (X) and Telegram, widely used for real-time news sharing, opinion formation, and group coordination, are particularly appealing

https://doi.org/10.1038/s41467-025-64785-1

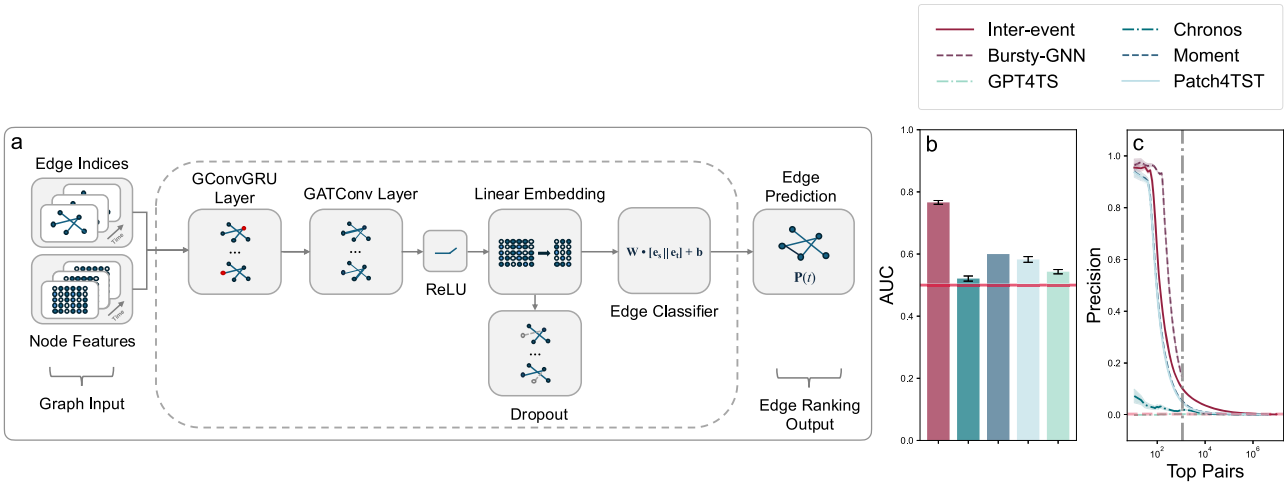

Fig. 5 | Scalability and the Bursty-TGNN. a presents the architecture of the 2-layer TGNN on top of temporal similarity networks. **b**, **c** present the average AUC and precision of the Bursty-TGNN, the vanilla bursty model and various temporal baseline models on the multi-domain identity matching problem, with the Bursty-TGNN manifesting an evident enhancement to the top-1000 precision. Error bars (**b**) and light shaded background (**c**) represent standard error.

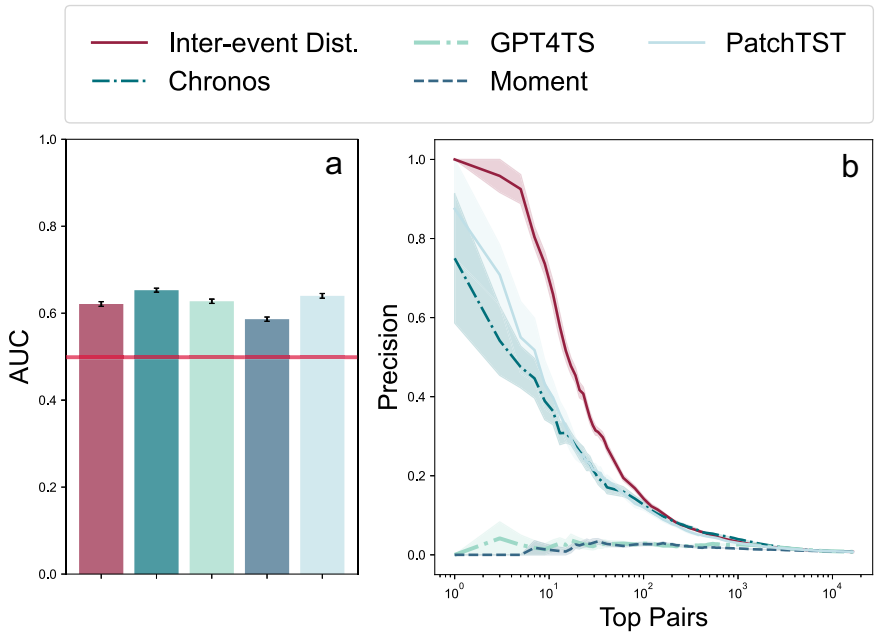

Fig. 6 | Performance evaluation of the inter-event similarity method (p^{ks}) for identity matching across social media platforms and comparison to temporal baselines. a depicts the averaged AUC over 8 weekly tests, with error bars signifying the standard error. **b** resents the average precision as a function of the examined number of pair candidates (with ±1 standard error in light background, correspondingly). The inter-event bursty model presents similar AUC and higher precision, comparing with baseline temporal models.

to actors seeking to amplify influence across diverse audiences. A critical step toward detecting such campaigns is the ability to accurately match identities across distinct platforms, especially when pseudonymity and content encryption hinder profile attribution across domains. In this experiment, we examine 266 profiles: 126 Twitter, 120 Telegram and 20 Instagram, corresponding to 131 different entities, on a weekly basis ($\Delta\tau$ = 7 days), for 8 weeks. Figure 6 presents the performance of the vanilla inter-event model and comparisons to the temporal baselines. The vanilla inter-event model outperforms the baseline models, attesting to the generalizability of the model to different platform types. To further evaluate generalizability, we conducted two additional identity matching experiments. The first experiment consists of matching profiles across different Reddit forums (sub-reddits), and the second one involved matching profiles across different Telegram channels. These analyses, presented in Supplementary Fig. 4, yielded higher performance than the Twitter-Telegram-Instagram setting and provide additional evidence for the generalizability of burstiness-based identity matching across diverse social media contexts.

Discussion

The widespread accessibility of generative AI and LLMs has made text generation remarkably fast and effortless, placing powerful tools within the reach of all. These technological advancements have led to a significant rise in manipulative cross-platform activity including the spread of sophisticated influence campaigns, which are now far more challenging to detect using content-based models due to their versatile and easily adaptable nature. Privacy restrictions, though essential

for safeguarding individuals, hinder the analysis of cross-domain patterns and correlations, which are crucial for identifying such coordinated activities. Countering these evolving threats across disconnected domains requires accurate identity linkage as a foundational step.

This study underscores the potential of temporal patterns, specifically individual burstiness, as a robust and scalable framework for cross-domain identity matching. We demonstrate that individual bursty dynamics form temporal fingerprints that persist across different platforms and over long periods of time, enabling accurate cross-domain identity matching. Our model outperforms state-of-the-art temporal and structure-based models in two experimental settings (Figs. 2 and 6), proving its generalizability. We demonstrated that the suggested model is stable over time, and is able to correctly identify 35% of the users even after an entire year, by examining at most 10 candidates for each profile (Fig. 3b). We further assessed the model's robustness to incomplete data and its impact on individual temporal fingerprints. We demonstrated that despite omitting various fractions of the individual's activity, temporal fingerprints remain similar, in particular for originally highly similar profiles (Fig. 4a, b). Although this affected the predictive ability of the model, it was still able to attain an AUC of 0.82 after an omission of 50% of individual activity (Fig. 4c). Lastly, we demonstrated that our model is highly scalable, outperforming the temporal baseline models in a setting involving 500 distinct domains (Fig. 5b, c), where all examined structure-based models failed to scale effectively. The limitations of structure-based models and the high computational complexity of existing temporal models highlight the advantages of our model as an efficient alternative, enabling real-time deployment on resource-limited devices.

Beyond the practical applications of cross-domain identity matching across encrypted domains, a deeper question emerges: why are temporal signals more informative for linking identities than structural ones? We postulate that structural patterns, influenced by connection type (e.g., trading, social, or professional networks), vary significantly across different platforms, obscuring structural coordination and hindering cross-domain identity matching. In contrast, the manifestation of temporal regularities stems from the interconnected nature of distinct domains. To explain this observation more generally, we turn to a theoretical perspective that captures how coordination can emerge across disconnected systems. Building on the "network of networks" framework[25–27], we consider coordinating individuals as bridges, implicitly linking seemingly disconnected domains. External events propagate as shock waves through this interconnected structure, while influencing coordinating entities across domains and triggering their actions. These actions, even if not simultaneous, often exhibit similar bursty patterns (see ref. 42 for formal modeling). Cross-domain identity matching, where a single entity controls multiple profiles across distinct platforms, offers a concrete example of this mechanism. The user's profiles implicitly connect otherwise disconnected domains: activity triggered by interaction or coordination on one platform can lead the same entity to act on another platform, thereby transmitting influence across systems, leading to similar bursty patterns. While previous studies primarily modeled human burstsiness based on isolated individuals, disregarding environmental effects[28,43–46], we offer a broader perspective, attributing aligned bursty behavior to shock waves traversing the network of networks.

These findings carry implications beyond the specific task of identity matching. In network science and computational social science, individual behavior is often modeled through structural relationships or semantic content. Our results suggest an alternative approach, modeling individuals based on the timing and dynamics of their actions. This temporal perspective enables modeling behavioral regularities across platforms, even in the absence of observable connections or shared metadata. By demonstrating that temporal signals are sufficient for identifying persistent behavior, our work positions time as a foundational dimension in understanding individual roles and coordinated activity within complex and fragmented systems.

Finally, it is worth considering whether coordination signals arising from bursty dynamics can be obscured. Our robustness analysis for the identity matching use-case (Fig. 4) reveals that despite omitting significant activity portions, most users remain identifiable. We hypothesize that this robustness may extend to the more general coordination scenario. Specifically, since many online settings are designated for timely responses, obscuring coordination from temporal signals is inherently challenging, as individuals naturally respond promptly to external shocks. Attackers may attempt to distribute actions over time using hidden agents, but such strategies are impractical. For instance, delayed transactions in money laundering raise suspicion, and dispersed actions weaken coordinated attacks' impact, suggesting that ultimately, attempts to obfuscate coordination may undermine the purpose of the coordinated activity itself.

Limitations and future research

An intrinsic limitation of our study is the need for sufficient individual data to reliably estimate inter-event distributions. Future research should examine how identification probability depends on activity volume and inspected period length. Our preliminary analysis demonstrates that identification probability increases with activity volume (Fig. S2) but decreases with longer inspection periods (Fig. S5), indicating short-term patterns are more effective for matching profiles. The choice of similarity measure also impacts performance, and alternative methods, such as those in ref. 47, could be evaluated. Further improving model scalability, possibly using complexity-reduced versions of the KS statistic[48] and optimizing search algorithms, is another promising direction. In addition, while our current approach is fully unsupervised, it would be valuable to explore the effect of incorporating limited supervision, such as fine-tuning with a small set of labeled identity pairs, to further enhance performance in low-resource scenarios. The authors intend to pursue these directions in future research.

Methods

Data

Financial markets data. We consider the Ethereum blockchain[35–37,49,50] as our financial dataset. This encrypted financial ecosystem enables the trading of tens of thousands of different crypto-tokens, using a single Ethereum wallet. Broadly, a crypto wallet is a digital tool that securely stores and manages the user's cryptocurrency holdings, allowing the user to send, receive, and monitor their digital assets on blockchain networks. The address of a crypto wallet serves as a unique identifier, similarly to an account number in traditional financial systems. Since a single Ethereum wallet can be employed for the trading of all Ethereum-based crypto-tokens, it can be used as the trader's identifier across different crypto-domains, for validation purposes (ground truth). We refer to a financial trading market D^i as encompassing all the trading activity related to the respective crypto-token c_i. We consider two different experimental settings over this dataset.

1. Two-domain setting: Considering 14 days of trading activity across two domains only, encompassing the activity of $2k$ daily users.
2. Multi-domain setting: Considering 14 days of trading activity across 512 financial domains, encompassing the activity of $250k$ daily users.

Both settings contain temporal data on an individual level granularity and network data, where an edge $(u, v) \in V_i$ represents that user u sold crypto-token i to user v.

Social platforms.

1. Cross Twitter-Telegram-Instagram: This dataset contains posting activity from 266 user profiles: 126 Twitter (X), 120 Telegram, and 20 Instagram, corresponding to 131 different individuals. The data was collected over a period of eight weeks (June 1 - July 28, 2024). We manually generated the ground truth labels for this dataset, relying on profile name similarity (for instance, "foo_networks" on Twitter and "FooNetworks" on Telegram), profile biographies, official websites, and other public sources of information (articles, Linktree) that indicate affiliation with the same individual or organization. The experiment on this dataset involved analyzing eight weekly activity snapshots of these users from all three social platforms. Each snapshot contains merely temporal data in the form of individual posting times, and lacks network data in the form of retweets, re-posts, likes and other connections.
2. Cross sub-Reddits: This dataset contains user-level temporal posting activity in multiple Reddit[51] forums (sub-Reddits), spanning over 3,075 users with verified activity in 2848 sub-reddits over a 19-week period. Ground truth for this dataset was defined based on shared Reddit accounts appearing in different sub-Reddit contexts. The experiment on this dataset entailed analyzing 19 weekly activity snapshots of these users on top of all active sub-Reddits. Each snapshot contains merely temporal data in the form of individual posting times, and lacks network data in the form of replies or mentions.
3. Cross Telegram Channels: This dataset contains posting activity from 248 Telegram users across 175 different channels, such that each user has posted in 2 or more of the channels. The data was collected over an 8-week period aggregated into weekly bins (Δt = 7 days). Ground truth for this dataset was defined based on shared Telegram accounts posting in different channels. Each of the eight snapshots contains merely temporal data in the form of individual posting times, and lacks network data in the form of replies or mentions.

Temporal graph neural network

We employ a temporal graph neural network (TGNN) on top of cross-domain similarity networks $G_{ks}^{\tau} = (V^{\tau}, E^{\tau})$, as described in section "Scalability: Bursty-GNN for temporal similarity networks." We frame this problem as a link prediction task, aiming to unveil potential connections within these similarity networks over time. Specifically, the task is identifying which profile pairs should be linked as corresponding to the same individual. This approach aligns with our objective to identify the implicit links between profiles, represented by edges in the similarity network, rather than to predict future transaction activity, as would be the case if we performed link prediction on the original transaction network. For each profile, we calculate 16 node features. Features include transaction counts, inter-event time statistics, and graph-based metrics like in-degree, out-degree, clustering coefficient, closeness, betweenness centrality, and PageRank. We use RobustScaler for feature normalization to mitigate the effects of outliers. Node features are recalculated daily to reflect the evolution of transaction behavior, providing a consistent set of features for training, validation, and evaluation.

Training. We employed supervised learning to train the TGNN, using edge labels informed by KS statistic measures. Positive training edges were defined as $KS \le \rho^p$, where $\rho^p = 0.001$, and negative training edges were defined as $KS \ge \rho^n$, where $\rho^n = 0.98$. A weighted binary cross-entropy loss function was used to counter class imbalance:

$$\mathcal{L}_{\mathrm{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[w_i \cdot y_i \cdot \log(\sigma(\mathrm{logits}_i)) + (1-y_i)\cdot \log(1-\sigma(\mathrm{logits}_i))\right] \tag{10}$$

where $\mathcal{L}_{\mathrm{BCE}}$ is the binary cross-entropy loss, N is the number of samples, w_i is the weight for sample i, y_i is the true label, σ is the sigmoid function, and logits_i are the raw scores from the classifier.

The Adam optimizer, initialized with a learning rate of 0.0001, was employed along with a ReduceLROnPlateau scheduler. Training was conducted for up to 100 epochs with early stopping to avoid overfitting. The data snapshots were divided into 80% for training and 20% for validation. A dropout rate of 0.5 was applied to the model. Each input node featured 34 attributes. The GConvGRU and GATConv layers were configured with 64 hidden units, and the output embedding dimension from the linear layer was set to 32.

Evaluation. We assessed the TGNN's capability to link Ethereum profiles against the baseline inter-event time distribution similarity. Consistent node features were used, and evaluation was performed across wallet activity for all days in the study period. The TGNN predicted the likelihood of an edge existing for each pair within two sets independently. First, previously seen edges, i.e., the 146 positive training edges. Then, the next 1000 most likely edges as ranked by KS score, corresponding to edges with a KS distance measure between 0.001 and 0.02616. The second set of edges had not previously been seen by the model.

Comparison baseline models

In this study, we compare the performance of the vanilla bursty model to various state-of-the-art models. We use each model in order to establish an embedding of the profiles, across the different domains. We apply cosine similarity on all embedding pairs to form a score, indicating the certainty that both profiles correspond to the same individual.

Structure-based models. We first compare against four baseline models, which rely on network data.

1. REGAL[14]: A spectral method solving the network alignment problem using network topology and nodes' feature similarity, followed by a low-rank matrix approximation speed-up.
2. IsoRank[15]: A spectral method propagating pairwise node similarity over the network employing the homophily assumption (profiles pertaining to the same individual across two domains would have similar topological network environments).
3. NAWAL[19]: Combines generative adversarial deep neural network with structural similarity to calculate node embeddings.
4. GAlign[20]: Calculates node embeddings using a multi-order Graph Convolutional Networks (GCN) to capture local and global structural information.

Temporal models. Foundation models are large-scale, pre-trained machine learning models, typically based on transformer architecture[52,53], designed for a wide range of downstream tasks across diverse domains. Building upon the success of foundation models in language and vision, foundation models for time series are large pre-trained models that capture complex patterns in temporal data across diverse domains and used for tasks like forecasting, sequence classification, anomaly detection, and imputation. In recent years, numerous foundation models for time-series analysis have been introduced[54], each employing unique methodologies to enhance forecasting accuracy and efficiency.

In this paper, we benchmark our bursty model against four SOTA foundation models for time-series analysis: CHRONOS, MOMENT, PatchTST, and GPT4TS. These models have been trained on datasets from various domains, spanning electricity, traffic, weather, and health, as well as the UCL and UEA collections. Given the scarcity of labeled data in many domains in time-series analysis, a particularly critical challenge for foundation models is zero-shot learning, which refers to the ability to perform tasks on unseen data without requiring

additional training or fine-tuning[55]. Benchmarking on zero-shot tasks evaluates the robustness and versatility of pre-trained models, highlighting their potential for real-world applications where task-specific data is limited or unavailable. Importantly, all models were reported to support zero-shot learning, and were evaluated on such settings.

1. GPT4TS[56]: Leverages pre-trained language models, specifically GPT-2, for time-series analysis. By treating time-series data similarly to textual data, GPT4TS applies the strengths of language models for forecasting tasks. The model fine-tunes the embedding layer, normalization layers, and output layer of GPT-2 to accommodate time series inputs, while keeping the self-attention and feedforward layers of the model unchanged. The model's short- and long-term forecasting performance has been tested on ETT, Weather, ILI, and ECL datasets[57] (Zhou et al.[56]; Ma et al.[41]), including zero-shot experiments on the ETT-H and -M datasets.
2. PatchTST[40]: A prominent transformer-based model tailored for long-term, multivariate time series forecasting. A key feature of the model is segmenting time-series data into subseries-level patches, which serve as input tokens to the transformer. This reduces the computational load by decreasing the length of input sequences and thereby lowering the time and space complexity associated with self-attention mechanisms. For zero-shot performance evaluation, it is tested on a diverse array of datasets spanning healthcare, finance and economics, retail and energy.
3. Chronos[39]: A family of pre-trained probabilistic time-series models that adapt language model architectures for time-series forecasting and encoding tasks, with parameter sizes ranging from 8 million (tiny) to 710 million (large). The model tokenizes time series data through scaling and quantization, enabling the application of language models to the data. The models are pre-trained on a collection of 13 datasets, encompassing energy, transportation, weather, and web traffic and have been evaluated on a wide range of datasets. For zero-shot performance, it is further tested on a wider array of datasets, including healthcare, retail, banking, and more.
4. Moment[38]: A family of foundation models designed for general-purpose time-series analysis. It addresses challenges in pre-training large models on time-series data by compiling a diverse collection of public datasets, termed the "Time Series Pile", which spans 5 large public databases. Moment employs multi-dataset pre-training to capture diverse time-series characteristics, enhancing its adaptability across various tasks such as forecasting, classification, anomaly detection, and imputation.

After extracting and grouping temporal information per user per time period, we compute embedding similarity between every pair of cross-platform users in the following manner: First, we align the temporal sequences with the desired model input length. This process includes left-padding or truncating time steps to match the desired model input (typically, $N = 512$ for small models and $N = 1024$ for large models). We then pass the padded tensor, adjusting dimensions as per model requirements, into the selected model in batches of 16 to accommodate GPU constraints. If applicable, we apply mean pooling to the output embeddings to achieve a one-dimensional vector representation per user. Computing the similarity is then straightforward: we apply cosine similarity between each pair of users, skipping self- and same-platform comparisons (as we are only interested in cross-platform identity matching) to derive a final similarity score.

Use of LLMs. The authors acknowledge the use of a LLM (ChatGPT-4o, OpenAI, San Francisco, CA) to improve grammar and enhance the clarity of the text. All AI-generated suggestions were critically reviewed and edited by the authors to ensure the original meaning was preserved. The authors take full responsibility for the content of the final manuscript.

Data availability

The financial markets data analyzed in this study is available from the dataset published in ref. 50, it has been deposited in the Harvard dataverse under accession code Ethereum-ERC20-markets. The Social media data analyzed in this study has been deposited in the posted on github alongside code for a minimal working example https://github.com/NetworkIntelligenceAndCoordinationLab/Latent-Connections-in-Social-Media.

Code availability

The code used for this study is available on github: https://github.com/NetworkIntelligenceAndCoordinationLab/Latent-Connections-in-Social-Media under MIT license.

References

1. Kasa, N., Dahbura, A., Ravoori, C. & Adams, S. Improving credit card fraud detection by profiling and clustering accounts. In *2019 Systems and Information Engineering Design Symposium (SIEDS)* 1–6 (IEEE, 2019).
2. Li, J., Wang, G. A. & Chen, H. Identity matching using personal and social identity features. *Inf. Syst. Front.* **13**, 101–113 (2011).
3. Viswanath, B. et al. Strength in numbers: robust tamper detection in crowd computations. In *Proc. 2015 ACM on Conference on Online Social Networks* 113–124 (ACM, 2015).
4. Kumar, S., Cheng, J., Leskovec, J. & Subrahmanian, V. An army of me: sockpuppets in online discussion communities. In *Proc. 26th International Conference on World Wide Web* 857–866 (ACM, 2017).
5. Wang, X., Peng, P., Wang, C. & Wang, G. You are your photographs: detecting multiple identities of vendors in the darknet marketplaces. In *Proc. 2018 on Asia Conference on Computer and Communications Security* 431–442 (ACM, 2018).
6. Tai, X. H., Soska, K. & Christin, N. Adversarial matching of dark net market vendor accounts. In *Proc. 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* 1871–1880 (ACM, 2019).
7. Zafarani, R. & Liu, H. Connecting users across social media sites: a behavioral-modeling approach. In *Proc. 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 41–49 (ACM, 2013).
8. Malhotra, A., Totti, L., Meira Jr, W., Kumaraguru, P. & Almeida, V. Studying user footprints in different online social networks. In *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* 1065–1070 (IEEE, 2012).
9. Goga, O., Perito, D., Lei, H., Teixeira, R. & Sommer, R. *Large-scale Correlation of Accounts Across Social Networks*. University of California at Berkeley, Berkeley, California, Tech. Rep. TR-13-002 (University of California at Berkeley, 2013).
10. Perito, D., Castelluccia, C., Kaafar, M. A. & Manils, P. How unique and traceable are usernames? In *International Symposium on Privacy Enhancing Technologies Symposium* 1–17 (Springer, 2011).
11. Liu, J. et al. What's in a name? An unsupervised approach to link users across communities. In *Proc. Sixth ACM International Conference on Web Search and Data Mining* 495–504 (ACM, 2013).
12. Mu, X. et al. User identity linkage by latent user space modelling. In *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 1775–1784 (ACM, 2016).
13. Labitzke, S., Taranu, I. & Hartenstein, H. What your friends tell others about you: low cost linkability of social network profiles. In *Proc. 5th International ACM Workshop on Social Network Mining and Analysis* 1065–1070 (ACM, 2011).
14. Heimann, M., Shen, H., Safavi, T. & Koutra, D. Regal: representation learning-based graph alignment. In *Proc. 27th ACM International Conference on Information And Knowledge Management* 117–126 (ACM, 2018).

15. Singh, R., Xu, J. & Berger, B. Global alignment of multiple protein interaction networks with application to functional orthology detection. *Proc. Natl. Acad. Sci. USA* **105**, 12763–12768 (2008).
16. Derr, T., Karimi, H., Liu, X., Xu, J. & Tang, J. Deep adversarial network alignment. In *Proc. 30th ACM International Conference on Information & Knowledge Management* 352–361 (ACM, 2021).
17. Nassar, H., Veldt, N., Mohammadi, S., Grama, A. & Gleich, D. F. Low rank spectral network alignment. In *Proc. 2018 World Wide Web Conference* 619–628 (ACM, 2018).
18. Crețu, A.-M. et al. Interaction data are identifiable even across long periods of time. *Nat. Commun.* **13**, 313 (2022).
19. Nguyen, T. T. et al. Structural representation learning for network alignment with self-supervised anchor links. *Expert Syst. Appl.* **165**, 113857 (2021).
20. Trung, H. T. et al. Adaptive network alignment with unsupervised and multi-order convolutional networks. In *2020 IEEE 36th International Conference on Data Engineering (ICDE)* 85–96 (IEEE, 2020).
21. Man, T., Shen, H., Liu, S., Jin, X. & Cheng, X. Predict anchor links across social networks via an embedding approach. In *International Joint Conference on Artificial Intelligence* 1823–1829 (ACM, 2016).
22. Zhou, F. et al. Deeplink: A deep learning approach for user identity linkage. In *IEEE INFOCOM 2018-IEEE Conference on Computer Communications* 1313–1321 (IEEE, 2018).
23. Chu, X. et al. Cross-network embedding for multi-network alignment. In *The World Wide Web Conference* 273–284 (ACM, 2019).
24. Riederer, C., Kim, Y., Chaintreau, A., Korula, N. & Lattanzi, S. Linking users across domains with location data: theory and validation. In *Proc. 25th International Conference on World Wide Web* 707–719 (ACM, 2016).
25. Kenett, D. Y. et al. in *Networks of Networks: The Last Frontier of Complexity* 3–36 (Springer, 2014).
26. Gao, J., Li, D. & Havlin, S. From a single network to a network of networks. *Natl. Sci. Rev.* **1**, 346–356 (2014).
27. Gao, J., Buldyrev, S. V., Havlin, S. & Stanley, H. E. Robustness of a network of networks. *Phys. Rev. Lett.* **107**, 195701 (2011).
28. Barabasi, A.-L. The origin of bursts and heavy tails in human dynamics. *Nature* **435**, 207–211 (2005).
29. Gonzalez, M. C., Hidalgo, C. A. & Barabasi, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).
30. Karsai, M. et al. Small but slow world: how network topology and burstiness slow down spreading. *Phys. Rev. E* **83**, 025102 (2011).
31. Oliveira, J. G. & Barabási, A.-L. Darwin and einstein correspondence patterns. *Nature* **437**, 1251–1251 (2005).
32. Zhou, T., Kiet, H. A.-T., Kim, B. J., Wang, B.-H. & Holme, P. Role of activity in human dynamics. *Europhys. Lett.* **82**, 28002 (2008).
33. Dezsö, Z. et al. Dynamics of information access on the web. *Phys. Rev. E* **73**, 066132 (2006).
34. Candia, J. et al. Uncovering individual and collective human dynamics from mobile phone records. *J. Phys. A Math. Theor.* **41**, 224015 (2008).
35. Buterin, V. A next-generation smart contract and decentralized application platform. *White Paper* **3**, 1–36 (2014).
36. Somin, S. et al. Remaining popular: power-law regularities in network dynamics. *EPJ Data Sci.* **11**, 61 (2022).
37. Somin, S., Altshuler, Y., 'Sandy'Pentland, A. & Shmueli, E. Beyond preferential attachment: falling of stars and survival of superstars. *R. Soc. Open Sci.* **9**, 220899 (2022).
38. Goswami, M. et al. MOMENT: a family of open time-series foundation models. In *Forty-first International Conference on Machine Learning* (ICML, 2024).
39. Ansari, A. F. et al. Chronos: learning the language of time series. In *Transactions on Machine Learning Research* (TMLR, 2024).
40. Nie, Y., Nguyen, N. H., Sinthong, P. & Kalagnanam, J. A time series is worth 64 words: long-term forecasting with transformers. In *The Eleventh International Conference on Learning Representations* (ICLR, 2023).
41. Jin, M. et al. Time-LLM: time series forecasting by reprogramming large language models. In *The Twelfth International Conference on Learning Representations* (ICLR, 2024).
42. Somin, S., Cohen, T., Kepner, J. & Pentland, A. Echoes of the hidden: uncovering coordination beyond network structure. Manuscript in preparation (2025).
43. Vázquez, A. et al. Modeling bursts and heavy tails in human dynamics. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **73**, 036127 (2006).
44. Malmgren, R. D., Stouffer, D. B., Campanharo, A. S. & Amaral, L. A. N. On universality in human correspondence activity. *Science* **325**, 1696–1700 (2009).
45. Malmgren, R. D., Stouffer, D. B., Motter, A. E. & Amaral, L. A. A poissonian explanation for heavy tails in e-mail communication. *Proc. Natl. Acad. Sci. USA* **105**, 18153–18158 (2008).
46. Gonçalves, B. & Ramasco, J. J. Human dynamics revealed through web analytics. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **78**, 026123 (2008).
47. Scholz, F. W. & Stephens, M. A. K-sample Anderson–Darling tests. *J. Am. Stat. Assoc.* **82**, 918–924 (1987).
48. Gonzalez, T., Sahni, S. & Franta, W. R. An efficient algorithm for the Kolmogorov-Smirnov and lilliefors tests. *ACM Trans. Math. Softw.* **3**, 60–64 (1977).
49. Somin, S., Altshuler, Y., Gordon, G., Pentland, A. & Shmueli, E. Network dynamics of a financial ecosystem. *Sci. Rep.* **10**, 4587 (2020).
50. Somin, S., Altshuler, Y. & Pentland, A. Crypto-asset trading on top of ethereum blockchain comprehensive dataset. *Sci. Data* **12**, 1407 (2025).
51. Baumgartner, J., Zannettou, S., Keegan, B., Squire, M. & Blackburn, J. The pushshift reddit dataset. In *Proc. International AAAI Conference on Web and Social Media* 830–839 (2020).
52. Raffel, C. et al. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.* **21**, 1–67 (2020).
53. Vaswani, A. Attention is all you need. In *Advances in Neural Information Processing Systems* (NIPS, 2017).
54. Liang, Y. et al. Foundation models for time series analysis: a tutorial and survey. In *Proc. 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* 6555–6565 (ACM, 2024).
55. Larochelle, H., Erhan, D. & Bengio, Y. Zero-data learning of new tasks. In *Proc. AAAI* 646–651 (ACM, 2008).
56. Zhou, T. et al. One fits all: power general time series analysis by pretrained LM. *Adv. Neural Inf. Process. Syst.* **36**, 43322–43355 (2023).
57. Sun, C., Li, H., Li, Y. & Hong, S. TEST: text prototype aligned embedding to activate LLM's ability for time series. In *The Twelfth International Conference on Learning Representations* (ICLR, 2024).

Acknowledgements

Research was sponsored by the United States Air Force Research Laboratory and the Department of the Air Force Artificial Intelligence Accelerator and was accomplished under Cooperative Agreement Number FA8750-19-2-1000 (A.P., J.K.). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Department of the Air Force or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

Author contributions

S.S. and K.E. designed the algorithm. S.S., K.E., and T.C. conducted the experiments. S.S. analyzed the results and wrote the manuscript with

input from all the authors. S.S., K.E., T.C., J.K., and A.P. reviewed the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-64785-1.

Correspondence and requests for materials should be addressed to Shahar Somin.

Peer review information *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.