TAPESTRY: Visualizing Interwoven Identities for Trust Provenance

Yifan Yang John Collomosse *
University of Surrey

Arthi Kanchana Manohar Jo Briggs Jamie Steane†
Northumbria University

ABSTRACT

In this paper we report our study involving an early prototype of TAPESTRY, a service to support people and businesses to connect safely online through the use of a Machine Learning generated visualization. Establishing the veracity of the person or business behind a pseudonomized identity, online, is a challenge for many people. In the burgeoning digital economy, finding ways to support good decision-making in potentially risky online exchanges is of vital importance. In this paper, we propose a Machine Learning method to extract temporal patterns from data on individuals' behavioral norms in their online activity. This monitors and communicates the coherence of these activities to others, especially those who are about to disclose personal information to the individual, in a visualization. We report findings from a user trial that examined how people accessed and interpreted the TAPESTRY visualization to inform their decisions on who to back in a mock crowdfunding campaign to evaluate its efficacy. The study proved the protocol of the Machine Learning method and qualitative insights are informing iterations of the visualization design to enhance user experience and support understanding.

Index Terms: [Machine Learning]: Topic modeling, Long Short Term Memory; [Human Computer Interaction]: Usability testing

1 INTRODUCTION

Personal privacy and trust online is a growing issue when we are increasingly reliant on systems involving disclosing potentially sensitive personal data. Online frauds, scams and other transgressions are escalating, in part as it is difficult to make good decisions about who to trust online. These are often facilitated by the fact that digital identities of people and services are presented through pseudonyms or anonymous addresses. How can we trust that the identity we are interacting with today wasn't created out of thin air yesterday to pull a scam; or, whether the service we are registering our personal data with, is trustworthy?

Traditional ways of authenticating online identities involve sharing static properties. These include entering personal details in an online form or using biometric features such as facial recognition. These properties are relatively easy to scrape, and thus plagiarize. This makes it difficult to discern the provenance and trustworthiness of the digital identity and compromises the security of online systems. Everyday, millions of people are challenged to make good decisions before disclosing their personal information.

Digital identity-related research increasingly involves exploring the behaviors and activities of an online user through the large volumes of data a user generates through e.g. social media posts and search histories that collectively comprise their digital footprint [5]. Such footprints play a crucial role across many digital economy services including user profiling [1], personality [7] and crowdfunding [10].

However, modeling individuals' user behavior norms in social media and other users' private data communicated online, as well as visually representing this to end-users, remains an open challenge that our research aims to address. In this paper, we conduct our study using aspects of individuals' digital footprints to model their digital identities to help determine and communicate the provenance of their online identity. Informing our research is our hypothesis that people have consistent (or slowly evolving) behavior and personal interests over longitudinal time periods [13, 14].

The paper goes on to report findings from a user study designed to test this 'identity provenance' as represented through visualization, in a prototype of TAPESTRY. We are developing the TAPESTRY platform, which stands for Trust, Authentication and Privacy over a DeCentralized Social Registry, within multidisciplinary research involving researchers from Machine Learning, Cryptography, Social Psychology and Design (Human Computer Interaction and communication design). We designed the reported user study in the context of a mock crowdfunding campaign to investigate how Machine Learning can examine and reveal historical digital footprint data as a behavioral norm, where deviation from this norm can be regarded as an anomaly, and where the coherence information extracted using our Machine Learning method generates the visualization (see Figure 2). In the study, the visualization aims to support users to make good decisions when pledging hypothetical money in the mock crowdfunding campaign. The campaign involves four real people from the gaming industry, who gave us their consent to use their public online data, alongside four faked personas. All eight crowdfunders are soliciting funds to develop a fictional new computer game called 'Earth Invaders'.

This paper makes the following contributions, with pertinence for the VizSec, InfoVIS and Human Computer Interaction communities in the first instance:

1) We contribute a Deep Learning method to extract temporal patterns of trust information using topic modeling;

2) Informed by this we report on our design and validation of a visualization to demonstrate the output of Deep Learning in a way that is accessible for users;

3) Following this we present findings from our user study involving the mock crowdfunding scenario and offer implications for trust extraction and visualization.

2 INTRODUCING TAPESTRY

The TAPESTRY platform proposes to collect, on an opt-in basis, digital trails generated by a subscriber's online interactions. Collectively, these comprise their digital footprint. These trails are left behind from everyday activities such as sharing photos, comments, 'likes' etc. from social media use, and also from an individual's search histories, interactions with IoT devices such as activity trackers, etc. One's digital footprint generates data that may then be used as a form of trust evidence that supports that identity [11].

In TAPESTRY, this trust evidence will be selectively encrypted and stored in a blockchain. These technologies are key to TAPESTRY's secure, privacy preserving function. Users will be able to grant third parties selective access to their own personal trust evidence, for a given time period, to prove the trustworthiness of their identity.

Meanwhile, from the online user perspective; a single visualization of the subscriber's multiple sources of trust evidence, across their different digital activities, is available on a dashboard. This pop-up dashboard forms the user interface part of the browser-based TAPESTRY tool. This aims to communicate between TAPESTRY

*e-mail: {yifan.yang, j.collomosse}@surrey.ac.uk
†email: {arthi.manohar, jo.briggs, jamie.steane}@northumbria.ac.uk

subscribers and website users the abstracted data in an accessible, comprehensible visualization. Meanwhile the personal privacy of the subscriber is retained as the specific sources of the data are not made explicit, and only the data relevant to a particular user context will be accessed on the blockchain. It is important to note here that TAPESTRY's aim is to support the user in making their own trust related decisions rather than to make decisions for them.

3 Visualizing Provenance

In this section, we explain the work-flow we adopted to process digital footprints to generate the visualization of the identity provenance. Specifically, users' updates posted on Twitter is the digital footprint source that we are interested in, informed by the study being a mock crowdfunding campaign. Twitter is a common place to share professional opinions, promote products and interact within peer communities. In this study, we used natural language processing methods to process textual information on our game developers Twitter feeds, and built a topic model based on this community. We then extracted temporal patterns learned using Deep Learning and compared the target pattern with reference pattern to detect incoherent temporal activity, which is regarded as one form of evidence for trustworthiness. Finally, we used Euclidean distance as a measurement, to represent the degree of incoherent activities on the game developers' time-lines using the proposed visualization.

3.1 Data and Preprocessing

The collection of data includes the identified crowdfunders (both real and fake profiles) and another 1632 anonymous game developer accounts. These game developer accounts are selected from the followers of a global game industry job board account (@GamesJobsDirect). Twitter time-line data was collected via the Twitter API[1]. Due to the API restrictions, we are limited to capture the latest 3200 tweets of a Twitter account. Retaining only the text of each tweet - discarding video or images - we then performed tokenization. We also removed stop words using NLTK[2].

3.2 Topical Word Modeling

A word embedding is a distributed representation of words, incorporating semantic information [6] that is learned from a large corpus of text (all tweets in the collected data set in our case). Topical modeling [1] extracts a distribution of words as topics, and a distribution of topics as documents. We implemented topical word embeddings, as proposed in [5], to capture contextual information in the given document. A topical word embedding is considered as a word-topic pair $< w_i, t_i >$. We considered all the tweets from one user as a document. The learned feature can enhance the discriminativeness among the words in different contexts and styles. A tweet embedding is the average of all topical word embeddings of the words in the tweet.

3.3 Temporal Coherence with Long-Short Term Memory

The application of Deep Learning [3] is proving highly effective in making sense of signals in computer vision [6], natural language [2] and robotics [9]. In this study, we apply Deep Learning to learn features for each individual game developer, denoted as user embedding. The user embedding is regarded as a temporal pattern of tweets in fixed time window (daily, monthly etc.).

The Long-Short Term Memory (LSTM) model [4] is a recurrent neural network used to model and predict time-series data. We built a sequence model to capture the coherence activities using a LSTM model and trained to extract the game developer's behavior norm based on their 'daily story', e.g. as played out on social media or through other online activity. We implemented a bidirectional LSTM to model the temporal coherence on a daily and weekly basis

[1]https://developer.twitter.com/en.html

[2]https://www.nltk.org/

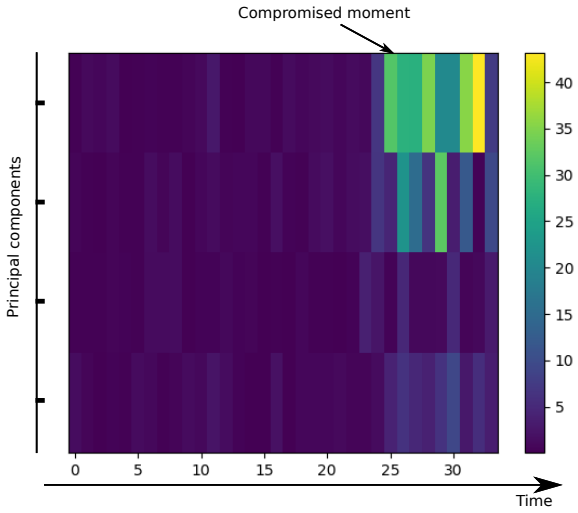

Figure 1: Hijacking detection visualization. The color of each block represents the similarity of the embedding of a temporal segment and reference in one of reduced principal components.

across the captured Twitter data (temporal segment). We adopted a two-layer bidirectional LSTM, followed by two fully connected layers. The input of LSTM is topical word embedding and the output is a daily or weekly tweets embedding.

We then applied supervised learning using Triplet Network strategy [12] for user embedding. The objective of this network structure is to put similar samples together and push dissimilar samples away from each other in embedding space. Here, similar samples are the temporal segments from one individual and dissimilar samples are the ones from different individuals. The method proved efficient in identifying different individuals from their temporal features, as learned with LSTM. We tested the method to detect compromised moments of an account, by randomly selecting a time step on the game developer's time-line feed. We then replaced the Tweets after the time point by the tweets from another game developer. The embedding of temporal segments were then projected in to a reduced dimension using Principal Component Analysis (PCA). In this experiment, we used the first 32 temporal segments (approximately one month's activity) as the game developer's reference (one game developer's behavior norm). We measured the distance between the target temporal segment and the average of reference segments in all principal components. Figure 1 is an example heatmap visualization for hijacking detection. The first two principal components in PCA reduced dimension can be used to discriminate between different individuals.

4 Visualization

Dealing with coherence of a user's behavior and making it easily understood, we translated the temporal patterns learned from the game developer's Twitter activities to a visualization output using a concentric circle design. Figure 2 shows early visualizations that have been generated through Machine Learning. To arrive at this design, the HCI researchers developed a number of possible visualization formats (see Figure 3) from which the communication designer went on to develop two. The main objective was to visually synthesize data gathered over the different digital activities and time periods while enabling relatively quick and easy comprehension. The inner two circles demonstrate daily tweet-related activities, and the outer two circles weekly activities. Tonal information meanwhile, similar to Figure 1, represents the coherence event on the time-line; the lighter tone shows coherent activities while the darker areas denote an absence of activities or activities that are incoherent with more usual ones.

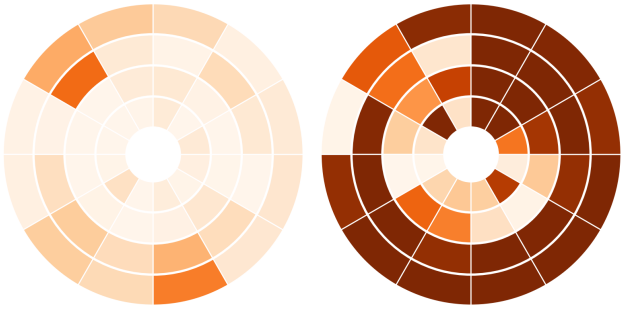

Figure 2: Visualization of two crowdfunders' digital footprints: strong provenance (left) and limited provenance (right). The two inner circles represent daily tweet activity and the two outer circles weekly.

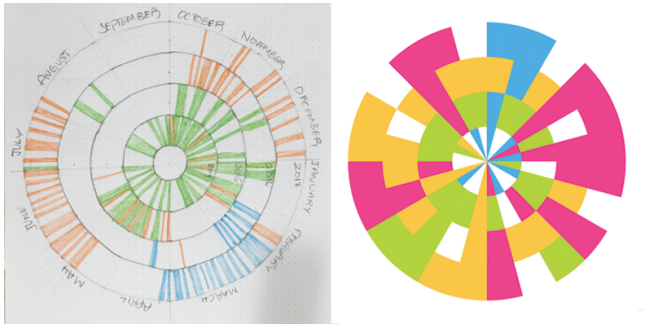

Figure 3: Early Communication Designs; concentric circles were used to represent the synthesis of temporal non-semantic information in an uncluttered universal form [8] for the early prototype trial.

5 User-study: crowdfunding campaign

Our study comprised a decision-making task experiment presented as a mock crowdfunding campaign where our four real gamers and four faked identities are pitching for hypothetical money. Through this trial, we analyzed the Machine Learning prototype and output visualization to test if our participants would make better investment decisions by selecting real gamers rather than frauds. We also commissioned an experienced video game story writer who mocked-up the new video game and crowdfunding campaign (Figure 4). We had gained the consent of the games developers to use their real profiles in the campaign. Meanwhile the writer produced four fake profiles based on their knowledge of the gaming industry. We created fake Twitter accounts for the fake profiles and continually tweeted relevant game and entrepreneur-related comments from March 2018. All eight crowdfunders had one campaign web page hosted on a password protected micro-site, which included a description of the game (constant across all candidates), a short biography for each profile and a link to their Twitter account. The creation date of the real and fake Twitter accounts was obfuscated.

5.1 Running the Study

We recruited 10 local university researchers to participate in the study in a computer lab. After a briefing on the TAPESTRY project, participants were invited to read the crowdfunding campaigns, browse the background description and biographies and invest a hypothetical $1000 'TAPESTRY currency' between the eight campaigns. We randomly split the participant group into two; one group was given the crowdfunding campaign pages. The second group was also provided with the TAPESTRY visualization as in Figure 2. Participants could use online sources that we provided or wider resources on the Internet to help them make decisions to allocate the money. The

Figure 4: The professional game writer created a Crowdfunding Campaign for the user studies; above shows the fake profile of Mark Stanton with the TAPESTRY visualization bottom right

study lasted 35 minutes; participants were asked to make one decision every 5 minutes given the knowledge they gleaned from their full use of Internet resources.

5.2 Results

We evaluated the participants' performance based on their investment results, comparing the amounts invested in real and fake profiles for both groups. From Figure 5, we can see that the accuracy of the investment results augmented along with the time. The more participants gathered information from their searches on the Internet, the more accurate they made their investment. Given the time limit, the TAPESTRY group used the visualization tool to quickly understand the games developers' Twitter identity, speeding up their search to establish legitimacy. We can conclude that although participants reached similar, correct decisions (in terms of discriminating their investment between genuine and fake developers) the time-to-task was considerable shorter (approximately by half) for TAPESTRY users. In the qualitative studies, issues that arose that we are now addressing include people's interpretation of lighter tone with less data (lighter Twitter use or, weaker provenance) rather than the converse. The selected orange hue was mis-perceived as associated with danger as in red for a warning. Otherwise, participants quickly interpreted the visualization and its purpose. Other participants were

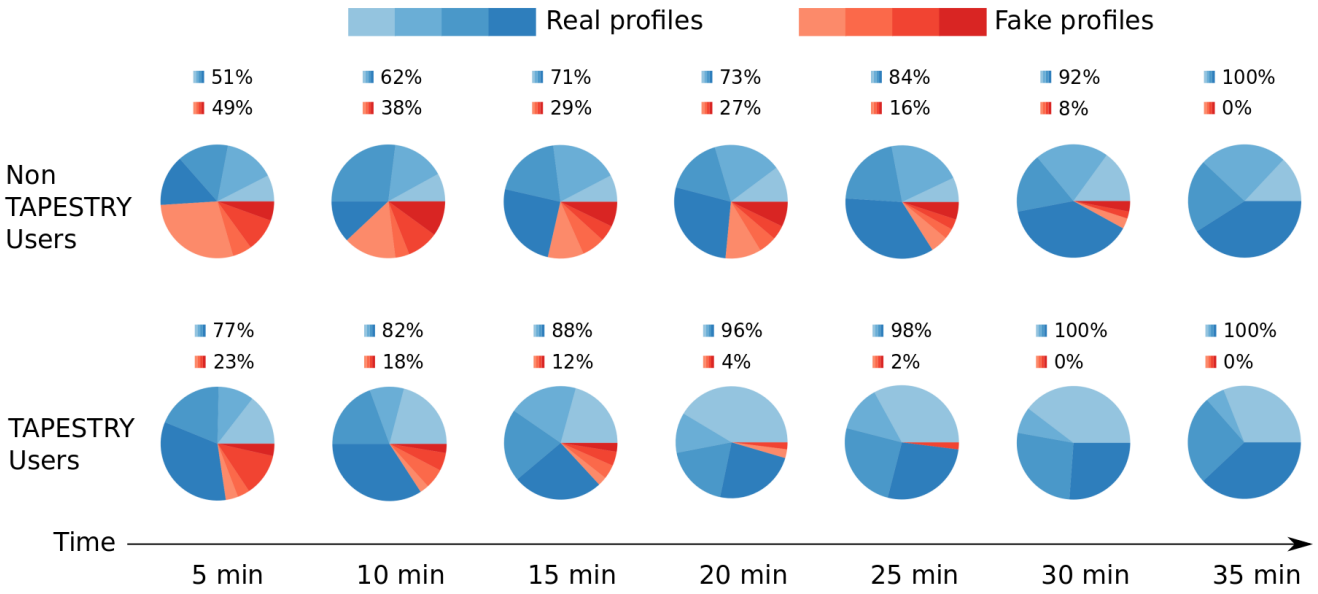

Figure 5: Participants' investment performance; top row is the group without TAPESTRY visualization; bottom row with TAPESTRY. The TAPESTRY users achieved better and faster decisions in this user study.

swayed by individuals' 'professionalism' signaled in links to positive media coverage or profiles on the sites of major gaming studios which had remained live from prior contracts. Conversely, most participants did not consider Twitter as a trusted source and thought it easily faked; as one participant put it: "*if it had been LinkedIn it would have been a lot easier [given the context of gaming and crowdfunding]!*" We aim to further investigate ways of identifying the main trusted sources, or more specially, key attributes of an online identity, which will need to be contextually relevant.

6 Conclusion

In this paper, we have reported our developed Machine Learning method to extract coherent activities on social media, as an evidence of the trustworthiness of identity provenance. On this basis, we created a visualization for end users to support them in making trust-related decisions within a mock crowdfunding campaign. The user study proved the protocol of the Machine Learning method based on topic models of game developers' behaviors in their Twitter accounts. In the next stage, cross-platform (e.g. LinkedIn, Facebook and web logs) and cross-modal signals (e.g. texts, photos and videos) processing is to be developed, as these forms of information can be cross validated amongst each other, to strengthen trustworthiness. The visualization was easily accessed and was interpreted by most of our study participants to quickly aid their decision-making tasks. We are currently iterating these with designers and conducting further trials to test users' understanding with a view to informing the final dashboard.

Acknowledgments

This work is funded by EPSRC project grant REF: EP/N02799X/1.

References

[1] G. Farnadi, J. Tang, M. De Cock, and M.-F. Moens. User profiling through deep multimodal fusion. In *Proceedings of the 11th ACM International Conference on Web Search and Data Mining*. ACM, 2018.

[2] Y. Goldberg. A primer on neural network models for natural language processing. *Journal of Artificial Intelligence Research*, 57:345–420, 2016.

[3] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. *Deep learning*, vol. 1. MIT press Cambridge, 2016.

[4] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[5] M. Kosinski, D. Stillwell, and T. Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, pp. 5802–5805, 2013.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[7] R. Lambiotte and M. Kosinski. Tracking the digital footprints of personality. *Proceedings of the IEEE*, 102(12):1934–1939, 2014.

[8] M. Lima. *The Book of Circles: Visualizing Sphere of Knowledge*. Princeton University Press, New York, 2017.

[9] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529, 2015.

[10] S. Nevin, R. Gleasure, P. O'Reilly, J. Feller, S. Li, and J. Cristoforo. Social identity and social media activities in equity crowdfunding. In *Proceedings of the 13th International Symposium on Open Collaboration*. ACM, 2017.

[11] L. Qiu, H. Lin, J. Ramsay, and F. Yang. You are what you tweet: Personality expression and perception on twitter. *Journal of Research in Personality*, 46(6):710–718, 2012.

[12] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815–823, 2015.

[13] B. Viswanath, M. A. Bashir, M. Crovella, S. Guha, K. P. Gummadi, B. Krishnamurthy, and A. Mislove. Towards detecting anomalous user behavior in online social networks. In *USENIX Security Symposium*, pp. 223–238, 2014.

[14] C. Wang and B. Yang. Composite behavioral modeling for identity theft detection in online social networks. *arXiv preprint arXiv:1801.06825*, 2018.