Mechanisms for Trust-Based Authentication of Fake Social Media Users

Alston Listenberger
Department of Computer Science
Purdue University
Indianapolis, USA
alistenb@iu.edu

***Abstract*—Social media has been consistently evolving to keep users engaged and participating for over a decade, increasingly growing the number of participants year after year who flock to social media to keep up with friends of get updated on the daily news. Since the beginning of the COVID-19 pandemic, more users are using social media than ever before and these users have spent more time on social media than in any year preceding the pandemic. This has led to a situation in which malicious actors attempt to capitalize on fears and anxiety in regards to the ongoing pandemic through the propagation of fraudulent information. With this rise of misinformation an urgent necessity has emerged to separate the carefully-crafted misinformation from the fact it is meant to obfuscate. This can be done through the application of varied trust-based mechanisms. However, the evaluation of trustworthy individuals in cyberspace can differ greatly from how one would evaluate an individual in person. The literature reviewed in this survey proposes a variety of methods to evaluate trust on social media.**

***Index Terms*—Trust, Social Media, Misinformation, Social Networks**

I. Introduction

Social Media is an incredibly powerful tool for humanity. Outside of giving us a pathway to keep in touch with friends and family, it has revolutionized the way in which we consume information. Where an event happening on the other side of the world would once be relegated to being reported on briefly during the evening news, we now can receive updates about any and every occurrence in real time, no matter where it has happened. In a sense, it has connected the world in a way that hasn't been dreamed in recorded history. In seconds, we can reach anyone anywhere in the world for a quick chat, but beyond this - within seconds - we can potentially reach millions of people through post sharing.

Like all tools, social media is also subject to unfortunate use cases ranging from benign to outright abuse. One such malicious use case is spreading carefully crafted misinformation. Since the beginning of the COVID-19 pandemic, more users are using social media than ever before and these users have spent more time on social media than in any year preceding the pandemic. These factors lay the groundwork to expertly cultivate elaborate misinformation campaigns that play on heightened fear and anxiety during uncertain times.

With more people than ever tuning into social media to acquire the pulse of the world, the conditions for false information to take root have been set meaning, due to the inherent dangers of the virus, misinformation on social networks is dangerous. So, what makes misinformation often able to take root in the place of certifiable fact? There are potentially a few key reasons. One being, not all information is equally valid. The best lies have a grain of truth in them, with the delineation between the two nearly imperceptible. This deceptive packaging of a falsehood inside of a fact makes accepting the falsehood palatable.

What is there to gain by propagating misinformation? Those with legitimate interest in spreading misinformation are likely spreading it for malicious reasons, under guise of a wholesome sharing of critical information. However, their intentions are far from wholesome. State actors could potentially leverage misinformation to instigate a restless and dissatisfied population in pursuit of global geopolitical gains, while a lone wolf actor could enjoy the thrill of misleading a multitude of people into dangerous situations.

While the reasons for misinformation propagating may vary, one thing is clear, trust is more crucial now than it has ever been when it comes to evaluating the information we receive. Trust is a familiar concept in basic human interaction. We are more likely to accept information from those we trust at face value and more likely to be skeptical of information when it is coming from a source we don't find trustworthy. In the past we primarily were concerned with determining how one might evaluate trustworthiness of acquaintances with whom we interacted with in person, but an increasingly important question is whether we can determine trust, in the sense of human interaction, virtually – and if so – how?

Trust in an online context is complex, thus, there are many proposed methods for handling this process. Pulling from the literature various methods were found such as detecting fake users, setting up online credibility sources, assigning users with a trust score, and ranked social recommendations. The aim of this survey will be to address this literature and generate a comprehensive estimation of how trust can be evaluated in the era of social media misinformation campaigns.

II. Detection of Fake User Accounts

The topic of misinformation propagation has been in the headlines at an alarming rate lately along with whistleblowers emerging to state social media companies aren't doing enough to prevent harmful information from spreading. With the changing false narratives being interwoven into articles of public health concern during the COVID-19 pandemic, one may wonder what more could be done to combat the emergence of this misinformation and who is spreading it. Often the origin of false information isn't a real user account, but a fake account created to appear legitimate. Other real users then spread the information which appeared to come from a legitimate source. Drawing inspiration from measurement theory's ability to measure human trust, Durresi et al. propose a trust-based security mechanism to identify clusters of fake user accounts on social media [13].

A. Methodology

The following section concerns methodological implementation of this framework.

1) Trust Measurements: Drawing inspiration from measurement theory, trust in human interaction can be quantified by a ratio of positive to negative interactions. Making a substitution for social media interaction, a proposed model of trust was generated based off of two factors: impression and confidence [13].

- Impression (m): Impression is the summation of one person's interactions with another. It is the aforementioned ratio of positive to negative interactions used to measure how much one individual trusts another individual. This equation for calculating impression is modeled in equation 1 [13].

$$m = \frac{\Sigma_{i=1}^{n} m_i}{n} \tag{1}$$

- Confidence (c): Confidence is a measurement concerning how certain a user is of their impression of the other user in question. This measurement is related to variance and is used to reduce the errors in measuring one's impression. Equation 2 shows the modeled equation for measuring confidence [13].

$$c = 1 - 2 \times v; where \text{ v} = \sqrt{\frac{\Sigma_{i=1}^{n}(m_i - m)^2}{n(n-1)}} \tag{2}$$

Of particular note, both of these values are utilized to measure trust and in the literature Duresi et al. make a note that in their calculations, these values were normalized to be between 0 and 1, but stated that the values can vary situationally [13].

While this is a solid foundation for calculating trust when there are direct interactions between users, indirect actions require a different approach. There are two methods listed in the literature: Trust Transitivity and Trust Aggregation [13].

- Trust Transitivity: This method is utilized to calculate indirect trust when an individual knows another user only through an intermediary (another user). For instance, it could be the observed interactions of a friend of a friend without direct interaction between the users. This measurement is a concatenation between the first user, the user's direct contact, and the indirect contact. The method of determining impression and confidence for this mode of indirect trust is modeled in equations 3 and 4 [13].

$$m_Z^{XY} = m^{XY} \oplus m^{YZ} \tag{3}$$

$$v_Z^{XY} = \sqrt{(v^{XY})^2(m^{YZ})^2 + (v^{YZ})^2(m^{XY})^2} \tag{4}$$

- Trust Aggregation: This method is utilized when there is more than one mode of indirect interaction between one user and another. Equations 5 and 6 show the modeled equations for measuring impression and confidence for this method of indirect trust measurement [13].

$$m_B^{AD} \oplus m_C^{AD} = \frac{m_B^{AD} + m_C^{AD}}{2} \tag{5}$$

$$v_B^{AD} \oplus v_C^{AD} = \sqrt{\frac{(v_B^{AD})^2 + (v_C^{AD})^2}{2^2}} \tag{6}$$

2) Clustering Graph: Social media functions as a network of people and groups, as such, undirected graphs G(V,E) can be generated with individual users as nodes (v) with the edges (E) being the relationship between any two given users. A spectral clustering method comprised of k-means clustering and DBSCAN is utilized to cluster large amounts of social media users into smaller clusters based on similarity [13].

3) Metrics: Some assumptions have been made to generate the metrics utilized here. A property assumed of fake users is that they tend to establish relationships between other fake user accounts, rarely branching out to establish connections with legitimate users. This could in part occur due to the inherent distrust of strangers attempting to add legitimate users. From this assumption, density emerges as a viable metric for cluster measurement and detection of fake user accounts. It is proposed that clusters of real users will be less dense than clusters of fake users. To this end, density is defined in equation 7 as [13]:

$$Density = \frac{2 \times \text{number of edges in graph}}{\text{Number of Vertices(Number of Vertices - 1)}} \tag{7}$$

However, there are considerations made for real users with a high-density cluster. In this instance, real users will have typically taken more time to build their cluster, whereas fake users will quickly add a large number of users, who also will have added many users in a short period of time. There are also trust-based factors to consider when evaluating clusters of users. Fake user clusters will have a high trust cluster as they need to build a perceived legitimacy so they are not easily identified as fake users, so they generate a lot of positive interactions with other fake users. Real users will have more variance in their interactions and, as such, will have varying degrees of trust amongst users they interact with [13].

B. Implementation

Fake social network graphs were generated using the Erdos-Renyi algorithm [13]. 1000 nodes were generated with some exhibiting good user behavior and some exhibiting high rate of connections to simulate fake users. Trust equations for impression and confidence were randomized for users, however these values were weighted to be higher for the simulated fake user accounts.

III. Establishing Online Credibility Sources

When one is attempting to ascertain the credibility of a website, one can look at its security certificate and verify that the information is accurate, aiding in the ability for a user to trust that a transaction is secure. If a certificate is fraudulent or missing, this could mean that a malicious third party is intercepting the traffic or that the entirety of the website itself is malicious. In this case, a user is notified by most modern web browsers and trust is likely not established. Such a system works extremely well, often in the background and is hardly noticeable by users at all unless there is a problem. These certificates are typically issued by a central, trusted authority, meaning that if there is a valid certificate, one can trust it. So, seemingly borrowing from this concept, Nunes and Correia have proposed a similar system for assigning online users a trustworthiness score in their paper, ”Improving Trust using Online Credibility Sources and Social Network Quality in P2P Marketplaces.”

A. Background

Peer-to-Peer (P2P) exchanges are exchanges in which a user exchanges something with another user. In the instance and model described in this paper, P2P exchanges are focused on commercial applications with users meeting online to buy or sell some form of commerce (buying, selling, trading, lending, renting) with the chance for a transfer of ownership occurring within some of these [12]. The authors note that while online marketplaces have rating methods for individual products, they do not currently have a method of rating individual buyers or sellers that can then be exported and applied elsewhere, making these ratings only meaningful in the context of the site they’re used on [12]. Extrapolating from this statement, the conclusion can be drawn that this leaves room for an individual to take advantage of users on one site until their rating has become poor and then leave for another site, beginning anew with a fresh ranking. The registration on most sites typically only requires that an email address be verifiable (in the form of clicking a link sent to the email to proceed with account activation). Everything else for these accounts typically can be forged as there is no true method for them to easily verify parameters such as if the name a user input is their real name, if the phone number they entered is linked to them or if it is just a temporary number, even if the other biographical information input by the user is factual. The authors note that these conditions ”foster a culture of anonymity [12].” It has been frequently found in literature that when people believe they are anonymous, they tend to behave differently, for example, because the potential for negative repercussions for their actions is perceived to be lower than if they are not anonymous. The authors propose that the best way to combat anonymity is to create trust online by delegating the certification of an individual’s trustworthiness to a central, ”trusted” authority [12].

B. Note on Methodology

The authors of this paper posit that the most widely adopted scale for measuring individual differences in trust, the Individual Trust Score, is dated and note that the current literature supports measuring trust in context-specific situations [12]. However, it is further noted that even the most recent proposed models of evaluating trust, for both theoretical and factorial models, fail to assign a practical scoring when attributing a trust score to a single user [12].

C. Methodology

While keeping the recent literature in mind, the authors address the deficiencies in trust measurements by proposing the WhyTrusted model. The system consists of three distinct functions [12]

- An ingestion engine
- A trust algorithm
- A profiling module

The ingestion engine focuses on aggregating public data from the website APIs and processing it. The trust algorithm operates in six steps [12]:

- Potential Trust Indicators (PTIs) are identified in the raw data passed in from the ingestion engine
- PTIs evaluated to determine if they can impact trust. If so they become Trust Var (TVar)
- Each TVar is normalized within the bounds of predefined values and weighted by the algorithm to calculate Trust Score (TS) and Source Rating (SR)
- Network Quality (NQ) is then evaluated further in SR calculation
- NQ of a particular network is done through a node evaluation of specific TVars output by that network’s API
- TS is finalized by incorporating the results of TVars, SRs, and NQ calculations

The authors note the formula is proprietary, so an evaluation of the actual algorithm cannot be conducted in the scope of this survey.

D. Results and Conclusions

The potential use for this model as well as limitations are identified in this section

1) Limitations in Methodology: The authors tested this model by running a survey of convenience, which typically doesn’t produce reliable results. The model was also tested by giving the survey respondents three simulated user trust scores in three categories

- High Trust Score
- Good Trust Score
- Poor Trust Score

with details of how the trust score was generated for each of the three users and asked the survey participants to choose which of the three user profiles seemed the most trustworthy. The ”high” and ”good” trust scores were heavily rated as the most trustworthy with ”better profile” and ”higher trust score” stated as the reasons for selecting these [12]. The flaw in this is telling the respondents that a user has a high trust score is likely to bias the results, with the participants likely opting to select a user they are told is trusted as being trustworthy. A different approach may be beneficial in order to determine if users would truly trust a score assigned by WhyTrusted.

2) Conclusions: While this model is designed to function cross-platform in P2P transaction situations, one can easily see where the application could be extended to the realm of social media ”transactions”, with the transactions being simply any interaction a user has with the social media site or other users of the social network. Any interaction a user has on a social media platform can increase or decrease the trust score depending on what action is taken. This method easily gives a social media platform the ability to identify and possibly handle harmful accounts on their network. Furthermore, in a similar concept as the P2P trust score being cross-platform, if multiple social media platforms implement such a trust score, this score could also easily be carried across the various social media platforms whom have chosen to implement the system, so factors which cause a lower ranking on one site would decrease a users' trust (and vice versa for increasing trust) across all sites implementing the scoring. This could be a very good and elegant solution towards efficiently stopping the spread of misinformation, as those who wish to spread it are likely to utilize multiple different sites in order to reach the widest possible audience. It would also allow trustworthy users to gain a great capability to combat this misinformation across platforms, even if their account is newly created on a platform by immediately being recognized as trusted users.

IV. Assigning Users with a Machine-Learning Generated 'Node Trust Score'

With the massive, ever increasing size of social networks, there is often more data than a human operator can handle. Machine Learning has increasingly been used to process this data and find meaningful connections, which are sometimes overlooked by human evaluators, in this aggregation of data. It is in aggregating and evaluating this data machine learning can be applied towards novel applications such as the generation of a user trust score for these sites.

A. Background

In this model, graph theory is applied to establish nodes and links (edges). The nodes are users of the social networking site who perform any action on the site. The links are the relationship between one user and another user who is an acquaintance of some form, whether it be a friend, family member, or friend of a friend to name a few [7]. With such a large number of users on these sites, it is likely a user will often encounter content from users which are not known to the user, and therefore the question of content trustworthiness becomes important. The goal of the authors is ”improve the quality of (content) recommendations and provide good user experience [7]” by proposing a model to ”evaluate trustworthiness for online social networks [7].”

B. Methodology

The authors generated a trust model utilizing three defined types of trust relationships between users and a social network [7]

- Inter-member trust (trust between members)
- User trust towards the provided service
- User trust towards the service provider

The authors have also developed a framework for calculating the node trust value using reinforcement learning and other machine learning methods. Different models are generated for different ”network topology [7].” The proposed framework by the authors is as follows [7]:

- Data is collected
- Features are selected relative to the social network
- A training set is constructed
- The training model is built
- The training model is used in the calculation of node trust values
- Simulation under real environment conditions

1) Feature Selection: In evaluating the trust value between two nodes, a and b, the following are defined by the authors. T(x) is the set of neighboring nodes relative to a given node, x [7]. K(x) is the number of neighboring nodes relative to a given node, x [7]. The utilized features are defined as follows [7]:

- Feature 1: Number of neighboring nodes, K(x)
- Feature 2: Sum and difference between number of neighboring nodes between two nodes, a and b. This is K(a) + K(b) and K(a) - K(b).
- Feature 3: Number of neighboring nodes in common, S(CN), between two given nodes, a and b. The formula for calculating this is as follows in equation 8:

  $$S(CN) = |T(a) \cap T(b)| \tag{8}$$

- Feature 4: How similar two given nodes, a and b, are. This is represented as follows in equation 9:

  $$Sim(a, b) \tag{9}$$

  Factors such as education level, geographic location, and personal interests [7] are included in this calculation. The authors note that the more similar two users are in these regards, the higher the likelihood they will establish trust.
- Feature 5: This feature is the Jaccard Index defined in equation 10:

  $$S_{ab}^{Jaccard} = \frac{|T(a) \cap T(b)|}{|T(a) \cup T(b)|} \tag{10}$$

  This is utilized as a method of normalizing node trust values in certain circumstances such as when two nodes both have a high number of neighboring nodes. In this instance, it may not necessarily mean these two nodes are very similar even though they share many common neighbors [7].

2) Training the Model: In the training set, the trust value of the nodes is replaced by the edge between the nodes [7]. If an edge is present, a 1 is assigned, else, a 0 is assigned as the trust value between the two given nodes [7]. The authors further limit the dataset by reasonably concluding that the further out in distance from the origin node, the less likely that there will be a connection between the origin node and another node. Therefore, the distance is limited to 3 [7]. The authors define the set, *f*, as the following equation represented in equation 11 [7]:

$$f = \{f_1, f_2, f_3, ..., f_n\} \tag{11}$$

The features can also be weighted as shown in equation 12 for logistic regression [7]:

$$\theta = \{\theta_1, \theta_2, ..., \theta_n\} \tag{12}$$

The logistic regression function when applied to the model is as defined in equation 13 [7]:

$$h_{\theta(f)} = g(\theta^T f) = g(\theta_1 f_1 + \theta_2 f_2 + ... + \theta_n f_n) = L \tag{13}$$

Where

$$g(x) = \frac{1}{1 + e^{-x}} \tag{14}$$

and L represents label information [7].The authors chose the logistic regression training model because the training speed is fast and the output, being a 0 or a 1, integrates nicely with node trust value scoring. The following process is used to train the model with logistic regression [7]:

- User relationships are constructed as a graph with distance between nodes ¡= 3. Then feature information for each node is calculated as described above.
- Set is trained using the logistic regression equations above to generate prediction model
- Prediction model is used to calculate trust value for the tested node pairs

C. Conclusions

The model appears to be sound. All of the data used can be acquired from public APIs on social networking sites. However, the description of feature 4 lacks a methodology as to how similar interests will be compared and identified, geographic location lacks descriptions as to what constitutes a similar enough geographic location and the same applies for education level, what would be considered a comparable level of education? Would it be the same (location independent) or within one degree of obtaining the same level of education? Or perhaps another arbitrary calculation for measuring likeness in education levels. As the algorithm relies on the identification of these features, it is critical to understand this factor to properly assess the model.

V. Modified Social Personalized Ranking Algorithm

A. Background

Social personalized ranking (SPR) algorithms are filtering algorithms which have been widely studied in academia and in the business sector. The authors have identified a deficiency in the literature in that prior studies have only ”integrated the user’s social network information into their model, without taking into account the transmission of social trust networks between users [10].” The authors note that the latest research on the subject shows that if these factors are taken into consideration, the algorithm can function better [10]. To improve the performance of SPR, the authors seek to incorporate the ”transitive social trust network between users into the SPR recommendation algorithms [10].”

B. Methodology

1) Mathematical Background: The authors make the following definitions [10]:

- A capital letter represents a matrix, for example: R
- R_{ij} represents an element in the matrix with i denoting the row and j denoting the column
- R^T is the transpose of the matrix
- Given $R = (R_{ij})_{m \times n}$, M represents the number of users and N is the number of items. R denotes an ’explicit feedback matrix [10]’ $R_{ij} \in \{0, 1, 2, 3, 4, 5\}$ or R_{ij} is unknown.
- $\hat{R}$ denotes a low rank matrix.
- The authors seek to approximate R with $\hat{R}$ where

$$\hat{R} = U^T V \tag{15}$$

$$U \in C^{K \times M} \tag{16}$$

$$V \in C^{K \times N} \tag{17}$$

- V and U are the explicit feature matrix for items and users
- $K << k$ and k is the rank of R
- $k \leq \min(M, N)$

2) The Model: The TrustSPR functionality is defined as follows:

$$\begin{aligned} &L(U, V) = \\ &\sum_{u=1}^{M} \{ - \sum_{i=1}^{N} I_{ui} \frac{exp(R_{ui})}{\sum_{k=1}^{N} I_{uk} exp(R_{uk})} log \frac{exp(g(U_u^T V_i))}{sum_{k=1}^{N} I_{uk} exp(g(U_u^T V_i))} \} \\ &+ \frac{\lambda_t}{2} \sum_{u=1}^{M} \sum_{v \in T(u)} (g(U_u^T W_v) - T_{uv})^2 \\ &+ \frac{\lambda}{2} (\sum_{u=1}^{M} ||U_u||_F^2 + \sum_{i=1}^{N} ||V_i||_F^2 + \sum_{v=1}^{M} ||W_v||_F^2) \end{aligned} \tag{18}$$

where

- $I(u)$ is representative of the set of all items that user u rated.
- $R(i)$ is the set of users who rated item i
- $T(u)$ is the set of users which user u trusts
- $T^+(v)$ is the set of all users who have trusted the user, v

C. Conclusions

The authors conclude, and present data, to show that this modification to traditional SPR algorithms improve performance, meaning this algorithm can likely be implemented as an effective way to filter out users which are not trustworthy on social media. Through the identification of users with many trusted others and others who trust them, a trust score could likely be collaborative. Perhaps an improvement could be that if a highly trusted user trusts another user, this could extend a greater benefit towards being trustworthy. Likewise for an untrusted user. As shown in Durest et. all [13] fraudulent users tend to group up initially and quickly build interactions amongst themselves. Perhaps this formula can build and quickly propagate a distrust score amongst these fake user clusters in order to better combat misinformation.

VI. Context-Aware Trust-based Matrix Factorization

Recommendation algorithms are a major part of modern living. Most systems in which a user interacts with now implement some form of recommendation algorithm, whether it be text recommendation for messaging, spell check, online search engines, even the ads we see on social media and video sharing sites are generated by recommendation algorithms. Li, Sun, and Lv propose a model trust recommender system in their paper, ”CTMF: Context-aware Trust-based Matrix Factorization With Implicit Trust Network [4].”

A. Background

The authors note that traditional recommender systems only consider the shallow relationship between a user and the item being recommended [4]. For instance, if a user is shopping online for clothing, ads can begin to recommend t-shirts to the user. However users tend to be looking for T-shirts in summer rather than winter, so the context is important. The authors also note that in real life, people tend to take recommendations only from those they trust and have a high degree of confidence in [4]. Therefore they propose a context-aware trust-based matrix factorization approach to focus on the implicit trust network built from user and item-based implicit trust [4].

B. Methodology

The authors of this paper propose two different constructions in order to improve ”improved trust-aware collaborative filtering” (ITACF) and incorporate these implicit trust network constructions into existing context-aware matrix factorization models with the goal of increasing rating prediction accuracy [4]. Trust propagation can be used to connect two unrelated users in a system by calculating their indirect local trust score utilizing the following process [4]:

- If there is one path between the two users, the minimum value of all direct trust scores between them is the indirect local trust score. The following formula can be used to calculate this score:

$$T_{u_a \to u_b} = MIN(T_{u_a \to u_{n1}}, T_{U_{n1} \to u_{n2}}, ..., T_{u_{ni} \to u_b}) \tag{19}$$

  where $T_{u_a \to u_b}$ is the local indirect trust score from user a, u_a to user b, u_b and u_{ni} are intermediate users in the pathing. The number of users cannot exceed the maximal degrees of separation, assigned as a maximal value for i.
- If multiple paths between two users exist, the indirect local trust score is calculated as the average value of all the indirect local trust scores from each path. The formula for calculating this metric is as follows [4]:

$$T_{u_a \to u_b} = AVG(T_{u_a \to u_b} P_1, T_{u_a \to u_b} P_2, ..., T_{u_a \to u_b} P_i) \tag{20}$$

  where $T_{u_a \to u_b} P_i$ represents the indirect local trust score from user a to b on path, i.
- Direct local trust and indirect local trust are combined into a final local trust score. Global trust is calculated utilizing a common method of global trust calculation [4].

If a user does not have implicit trust neighbors, then the model is limited. The authors construct an item-based implicit trust model to address this [4].

C. Conclusions

The authors conclude that they have made four improvements to the existing context-aware matrix factorization approaches to recommender systems [4]:

- Trust propagation is used to calculate a local trust score from null, or one for a new user
- A global trust score is calculated using the local trust score
- Item-based implicit trust is considered
- Different trust networks are able to be constructed contextually, in different conditions

They continue to note that these improvements solve the following problems [4]:

- The problem of the initial rating matrix being too sparse
- The problem of malicious attack on the rating system
- The problem for users who ”cold start” or just begin an account on a new network

The approach the authors of this paper use incorporates improvements to existing algorithms which have been identified in literature as weaknesses, however the effectiveness of the algorithm in recommending still needs to be compared with other similar algorithms to determine relative effectiveness. I believe that addressing the above problems can lead to big strides in combating misinformation before it begins as these users often create new accounts and attempt to build trust quickly in order to begin spreading effective misinformation in which other, unwitting users, are tricked into believing as factual.

VII. Conclusions

COVID-19 disinformation on social media has created a dangerous environment where the problem of misinformation, once relegated to simply being annoying in online circles, has become dangerous. Harmful medical disinformation, relevant to the most vulnerable parts of the population, is being propagated at an alarming rate and helpful medical information is being covered up and made difficult to find. Exacerbating the problem, the most vulnerable population towards the effects of the virus, the elderly, are also the most prone to believing misinformation online. A possible solution to this problem is to build trust frameworks into the very structure of the web so misinformation could be combated before it begins to propagate harm. After surveying these papers, I can conclude that it is highly likely a combinatorial approach utilizing pieces of the above methods could be the most beneficial. Some identified weaknesses of papers such as newer users not being trusted or distrusted inherently are addressed by solutions in other papers such as the ”cold start” solution in ”Context-Aware Trust-Based Matrix Factorization.” A central authority as proposed in ”Establishing Online Credibility Sources” could provide a basis for transferring scores to networks in which a ”cold start” solution isn’t applicable. By generating truly globalized trust scores, misinformation could be weakened, and perhaps, defeated on the internet. Modified social personalized ranking algorithms could potentially be utilized to recommend factual news and medical information to users or even highly trusted individuals as connections, while filtering out individuals which are not trustworthy. The detection of these fraudulent users could be done utilizing the clustering algorithms in ”Detection of Fake User Accounts” while the machine learning algorithms from ”Assigning Users with a Machine-Learning Generated ‘Node Trust Score’” generate and assign trust scores to these accounts. More research on how to best incorporate the findings in this survey together is required, but is beyond the scope of this paper.

Acknowledgment

This paper was prepared as a course capstone for CSCI 57500/49000 in the Department of Computer Science at Indiana University/Purdue University Indianapolis. While great care was taken to ensure accuracy of surveyed papers and cited references, if a mistake is noted, please feel free to contact the author for correction.

References

The following references were read and utilized in the writing of this survey paper

References

[1] Y. Wang, J. Wen, W. Zhou and F. Luo, ”A Novel Dynamic Cloud Service Trust Evaluation Model in Cloud Computing,” 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), 2018, pp. 10-15, doi: 10.1109/TrustCom/BigDataSE.2018.00012

[2] S. Tucker, ”Engineering Trust: A Graph-Based Algorithm for Modeling, Validating, and Evaluating Trust,” 2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE), 2018, pp. 1-9, doi: 10.1109/TrustCom/BigDataSE.2018.00011

[3] C. Li, R. Li, L. Zhuang and X. Zhang, ”Formal Analysis of Trust Chain,” 2010 Second International Conference on Networks Security, Wireless Communications and Trusted Computing, 2010, pp. 111-116, doi: 10.1109/NSWCTC.2010.34

[4] J. Li, C. Sun and J. Lv, ”CTMF: Context-Aware Trust-Based Matrix Factorization with Implicit Trust Network,” 2014 Seventh International Symposium on Computational Intelligence and Design, 2014, pp. 387-390, doi: 10.1109/ISCID.2014.176

[5] Z. Lin and L. Dong, ”Clarifying Trust in Social Internet of Things (Extended Abstract),” 2018 IEEE 34th International Conference on Data Engineering (ICDE), 2018, pp. 1825-1826, doi: 10.1109/ICDE.2018.00270.

[6] F. Hao, S. S. Yau, G. Min and L. T. Yang, ”Detecting k-Balanced Trusted Cliques in Signed Social Networks,” in IEEE Internet Computing, vol. 18, no. 2, pp. 24-31, Mar.-Apr. 2014, doi: 10.1109/MIC.2014.25.

[7] H. Mayadunna and L. Rupasinghe, ”A Trust Evaluation Model for Online Social Networks,” 2018 National Information Technology Conference (NITC), 2018, pp. 1-6, doi: 10.1109/NITC.2018.8550080.

[8] W. Chang and A. N. Diaz, ”How Can Social Networks Help Us Measure Trust Online?,” 2012 Ninth International Conference on Information Technology - New Generations, 2012, pp. 865-866, doi: 10.1109/ITNG.2012.76

[9] W. Yuji, ”The Trust Value Calculating for Social Network Based on Machine Learning,” 2017 9th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2017, pp. 133-136, doi: 10.1109/IHMSC.2017.145

[10] G. Li, Y. Chen, Z. Zhang, J. Zhong and W. Ou, ”Social personalized ranking recommendation algorithm by trust,” 2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), 2017, pp. 273-277, doi: 10.1109/SPAC.2017.8304289.

[11] G. Bachi, M. Coscia, A. Monreale and F. Giannotti, ”Classifying Trust/Distrust Relationships in Online Social Networks,” 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing, 2012, pp. 552-557, doi: 10.1109/SocialCom-PASSAT.2012.115

[12] M. Nunes and J. Correia, ”Improving trust using online credibility sources and social network quality in P2P marketplaces,” 2013 8th Iberian Conference on Information Systems and Technologies (CISTI), 2013, pp. 1-4.

[13] Kaur D., Uslu S., Durresi A. (2019) Trust-Based Security Mechanism for Detecting Clusters of Fake Users in Social Networks. In: Barolli L., Takizawa M., Xhafa F., Enokido T. (eds) Web, Artificial Intelligence and Network Applications. WAINA 2019. Advances in Intelligent Systems and Computing, vol 927. Springer, Cham. https://doi.org/10.1007/978-3-030-15035-8-62