Trust and Trustworthiness in Social Recommender Systems

Taha Hassan
Computer Science Department, Virginia Tech
Blacksburg, VA
taha@vt.edu

D. Scott McCrickard
Computer Science Department, Virginia Tech
Blacksburg, VA
mccricks@vt.edu

ABSTRACT

The prevalence of misinformation on online social media has tangible empirical connections to increasing political polarization and partisan antipathy in the United States. Ranking algorithms for social recommendation often encode broad assumptions about network structure (like homophily) and group cognition (like, social action is largely imitative). Assumptions like these can be naïve and exclusionary in the era of fake news and ideological uniformity towards the political poles. We examine these assumptions with aid from the user-centric framework of *trustworthiness* in social recommendation. The constituent dimensions of trustworthiness (diversity, transparency, explainability, disruption) highlight new opportunities for discouraging dogmatization and building decision-aware, transparent news recommender systems.

KEYWORDS

Recommender systems, social recommendation, collaborative filtering, trust, transparency, explainability, diversification, quality

1 INTRODUCTION

In the 2016 US Presidential Election, social media was one of the key venues for dissemination of fake news, with aggregate viewership of fake news strongly correlated with aggregate voting patterns [3]. These patterns hint at deeper geographical and sociopolitical groupings in the US (see figure 1). Differences in media habits by political affiliation have also been the subject of attention of late. A Pew Research Center study [1] found little evidence of overlap between news sources frequented and trusted by self-identified consistent liberals and conservatives. Conservatives gravitated towards a small number of news outlets, often just one, and exhibited distrust of a majority of the mainstream news sources. Liberals consulted a wider variety of news outlets, with social issue-based choices. In addition, consistent liberals were more likely to 'unfriend' or block someone because of their political views, while consistent conservatives were more likely to have friends who agree with their political views. According to the Pew Center study, Facebook draws twice
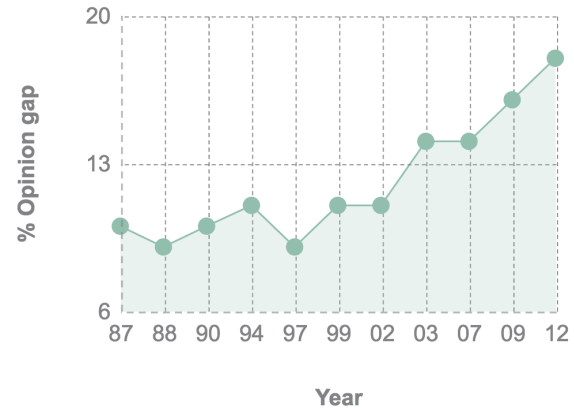
Figure 1: Partisan differences in political values (average percentage difference between Republicans and Democrats) on a 40-point values questionnaire [1]
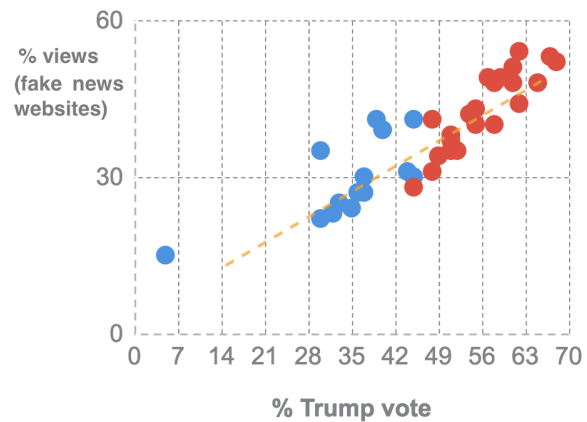

Figure 2: Voting behavior (by state) and average daily viewership of visitors to websites serving fake news. Blue represents a state-wide Democratic victory, red represents a Republican victory [3]

as many consumers of news on politics and government relative to Yahoo or Google News, and about 40% of the US population would seek these news stories on Facebook week before the survey. A number of social recommender systems serving content as varied as pages, groups, events and jobs, work behind-the-scenes at Facebook. User interaction with a social recommender encodes interesting technical, behavioral and demographic trends, especially in the era of increasing partisan rifts. This study explores key limitations of

social recommendation from a critical-theoretic standpoint and provides new design considerations for their future.

The next sections review the conventional notion of trust-awareness, present a critique of this notion within the context of political polarization and explore issues of transparency, explainability and diversity for social news recommenders.

2 RELATED WORK

Implementations of recommender systems have been traditionally attributed to one of two approaches: the model-driven and the data-driven [8]. The first, known as collaborative filtering (CF) refers to modeling the users and items jointly, by a shared space of latent feature vectors [4]. The second, neighborhood modeling (NM), identifies user-user and item-item similarities directly. In practice, these two approaches have similar ancestry and can be put together in hybrid approaches.

Trust-based recommender systems are concerned with learning the preferences of trustworthy social neighbors, or 'friends' of an individual user, as well as the mistrusted 'foes' [12]. These preferences inform the latent features inferred in CF, such that features for an individual are ranked closer to his or her friends' features, rather than the foes. In [9], learning-to-rank models minimize a loss function on a personalized ranking function. Using trust and mistrust relationships proves effective in combating the sparsity inherent in users' preferences data.

User-perceived quality of recommender systems' output is a broad way to evaluate anything from aggregate emotional impact to perceived relevance and variety. User-perceived variety or diversity in the recommender output is a related, important notion. Authors in [5] explore an organization interface (ORG) as opposed to a list interface for a top-N style recommender. This interface clusters and annotates subgroups of the recommender output. These annotations go beyond category labels and express tradeoffs in product quality and price. ORG ranked better than lists in perceived ease of use and diversity. Studies like [15], however, note that statistical accuracy of such recommendations might be lower, even as they are rated better in quality by field trial participants.

3 TRUSTWORTHINESS

3.1 A Critique of Trust

User feedback in recommenders, be it explicit (ratings, reviews) or implicit (click behavior, browsing habits) is often sparse and non-generalizable. This sparsity has driven the need for the conventional notion of trust. Often leveraged in CF, this notion of trust assumes homophily in the social network, that like connects with or favors like [3]. Trust-aware CF, therefore, ranks the judgment of trustworthy friends as consistently higher than those of untrustworthy peers and foes. From a broad critical-theoretic standpoint, this construction of trust can be arbitrary and cynical. Note that sparsity encodes an emergent market need for showing relevant news stories rather than in-depth or disruptive ones. Also note that, the qualifier 'emergent' helps avoid the trendy critique of reverse-cynicism. It refers to the aggregate of choices by a large number of people potentially exercising their free will, choices such as reviewing or rating items less frequently than idly browsing a news feed.
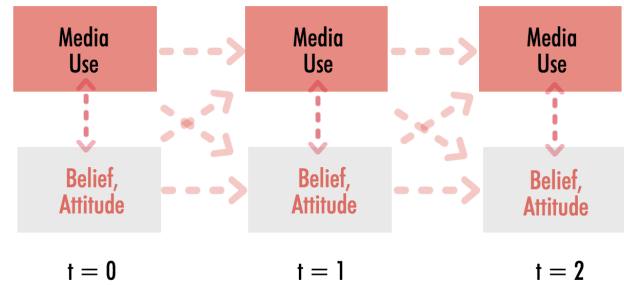

Figure 3: A three-way, lagged model of behavior reinforcement spirals [13].

To expand this critique, we can turn to a large volume of literature in the behavioral sciences domain [14][13] that examines the degree and scope of recommendation influence on online choices. Recall how consistent conservatives and liberals have strong selective tendencies in choosing whom to be friends with and what media sources to trust. The author in [13] presents a theoretical framework for bias reinforcement in social media (see figure 3). A closed feedback loop aims to maximize the use of a single social medium in the absence of competing interests. While a statistical analysis of these models is beyond the scope of the study, there are important conceptual aspects to consider here. One, consider the multitude of paths that can lead to feedback loops of gratification. Two, consider the assumptions behind nominal categories like media use, such as that of a closed system. In practice, institutions like marriage, family and religion, as well as competing media sources might help limit the scope of this reinforcement.

While there is some prior work on longitudinal trends in recommender output and rating diversity [10], reinforcement loops have not been formally studied from a recommender UX perspective (to the best of the author's knowledge). It is helpful, therefore, to reflect on the interactions that facilitate such loops. Authors in [7] describe a user-centric framework for perceived effort, accuracy and satisfaction of a recommender systems user (see figure 4). A certain dual causation exists between *experience* and *interaction* in this UX model. *Experience* can inform and modify *interaction*, as much as *interaction* informs and modifies *experience*. We can thus, leverage the spiral model in [13] to conceptualize bias-reinforcement as an undesirable byproduct of this duality. Furthermore, the study reports that the users who rate the recommendations as 'effective' have on average, a lower total viewing time and total number of watched clips. They do not necessarily consume a lesser volume of information, but less of their time is spent browsing than it is on watching recommended clips.

Another important distinction is the so-called 'greedy' approach to search. Some users are looking to find the 'best' possible recommendation while browsing, while others are looking to find the first perceived optimal recommendation. Domain expertise and trust affect these decisions [6]. In that, the tendency to favor sophisticated control (like, user-defined weights and order) over sort-only control correlates well with domain expertise. An 'average' user is likely a lot more passive in taking a recommender system's output at face value. Authors in [10] report that for long-term users of the
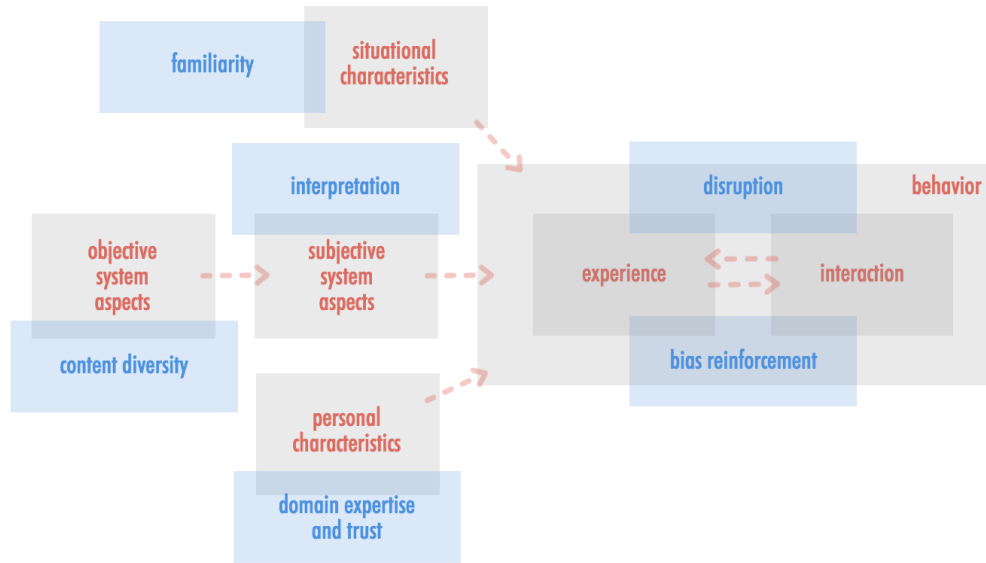
Figure 4: A recommender systems UX framework [7] from the perspective of trust and disruption.

MovieLens recommender, the diversity of recommender output and of explicit user ratings narrows over time. However, the drop in diversity is greater for users who tend to ignore the recommender output whilst rating content relative to those who do not.

This calls for design considerations that go beyond the local-neighborhood approach to recommendations. The output of a recommender system should help disrupt rather than reinforce feedback loops [2]. This involves a move towards explainability and transparency, as well as diversification and provocation.

3.2 Conceptual Dimensions of Trustworthiness

In the broadest sense, a trustworthy recommender is a recommender that enables transparent and interpretable interaction against the rising current of dogmatization and partisan antipathy. The key conceptual dimensions are as follows.

3.2.1 Diversification. Diversification for a news recommender improves the potential for a given user's encounter with alternative perspectives on a given social issue and disrupts the feedback loops explored in the previous section. There is room for creating a notion of diversity using a quality ranking for each news article. Google News, for instance, features 'in-depth' articles for a subset of search queries based on automated strategies. Similarly, Amazon.com recommenders arrange reviews by usefulness based in part on explicit feedback by readers of said reviews. A diversified recommender should recognize the degree of substantiation, i.e., it should be able to distinguish between an 'in-depth' analysis or news story versus a 'broad overview' and present a mix of labeled samples from each. Empirical studies in [6] and [7] associate the tendency to tag with a user's trust of the recommendation system and lack of privacy concerns. Therefore, manual tagging of article quality

needs to be evaluated for agreement between independent labelers and between labelers and third-party fact-checking agencies.

3.2.2 Disruption. Model-level disruption is a more intrusive strategy that can potentially challenge the neighborhood assumption of recommender systems and create more provocative collections of news and analyses. The present state of recommender systems makes it difficult to get recommendations on a specified social perspective. In other words, we cannot ask a recommender system to fetch say an anarcho-syndicalist or a second-wave feminist perspective on a social issue or a news story, unless there is explicit or implicit history of a user browsing articles similar to such perspectives. However, consider the notion of transfer learning. Transfer learning for CF [11] allows cross-domain application of machine-learnt models and sparsity reduction. There is, however, no transfer-learning framework that uses a user-specified target domain to rank and retrieve recommendations, to the best of the author's knowledge.

3.2.3 Explainability and Interpretation. Another facet of explainability (as well as transparency) is decision-awareness for the user. Netflix's movie recommender, for instance, includes annotations on why it ranked a given movie or a TV show highly among the recommended content. These include references to actors, directors, genres and relevant decades, among others. A news recommender, especially one concerned with popular conversations, can potentially increase the users' trust and confidence in its output using explanatory annotations. This might especially be true with automated flagging of ad spam, content involving bullying and online predatory behavior, or NSFW images and video data.

4 FUTURE WORK

This study explores the theoretical foundations of trust-aware ranking in social recommenders. A key frontier in building such systems is verifying the user-perceived quality of disruptive interaction. Therefore, crowd-sourced tagging of article quality and authenticity, third-party verification, progressive disclosure of perspectives in a conversation and disruptive personas are subject of future empirical investigations by the author.

5 CONCLUSION

Disruption in recommender systems can take one of many forms. However, the relatively new interest in this domain implies a lack of frameworks, empirical studies, field trials and ethnographic work. Culture-awareness is an ambitious frontier for recommender systems of the future, and one that encompasses ideas like model differentiation, language preservation and community building. Diversification, transparency and disruption are key building blocks for tools that enable these ideas.

REFERENCES

[1] Jocelyn Kiley Amy Mitchell, Jeffrey Gottfried and Katrina Eva Matsa. 2014. Pew Research Center: Political polarization and media habits. (2014). Retrieved 2017 from http://www.journalism.org/2014/10/21/political-polarization-media-habits/
[2] Shaowen Bardzell, Jeffrey Bardzell, Jodi Forlizzi, John Zimmerman, and John Antanitis. 2012. Critical design and critical theory: the challenge of designing for provocation. In *Proceedings of the Designing Interactive Systems Conference.* ACM, 288–297.
[3] Adam Fourney, Miklos Z Racz, Gireeja Ranade, Markus Mobius, and Eric Horvitz. 2017. Geographic and Temporal Trends in Fake News Consumption During the 2016 US Presidential Election. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management.* ACM, 2071–2074.
[4] Taha Hasan, Naveed Arshad, Erik Dahlquist, and Scott McCrickard. 2017. Collaborative filtering for household load prediction given contextual information. In *Proceedings of the 2017 SIAM Workshop on Machine Learning for Recommender Systems (MLRec).*
[5] Rong Hu and Pearl Pu. 2011. Enhancing recommendation diversity with organization interfaces. In *Proceedings of the 16th international conference on Intelligent user interfaces.* ACM, 347–350.
[6] Bart P Knijnenburg, Niels JM Reijmer, and Martijn C Willemsen. 2011. Each to his own: how different users call for different interaction methods in recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems.* ACM, 141–148.
[7] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction* 22, 4-5 (2012), 441–504.
[8] Yehuda Koren and Robert Bell. 2015. Advances in collaborative filtering. In *Recommender systems handbook.* Springer, 77–118.
[9] Paolo Massa and Paolo Avesani. 2007. Trust-aware recommender systems. In *Proceedings of the 2007 ACM conference on Recommender systems.* ACM, 17–24.
[10] Tien T Nguyen, Pik-Mai Hui, F Maxwell Harper, Loren Terveen, and Joseph A Konstan. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd international conference on World wide web.* ACM, 677–686.
[11] Weike Pan, Evan Wei Xiang, Nathan Nan Liu, and Qiang Yang. 2010. Transfer Learning in Collaborative Filtering for Sparsity Reduction.. In *AAAI*, Vol. 10. 230–235.
[12] Dimitrios Rafailidis and Fabio Crestani. 2017. Learning to Rank with Trust and Distrust in Recommender Systems. In *Proceedings of the Eleventh ACM Conference on Recommender Systems.* ACM, 5–13.
[13] Michael D Slater. 2007. Reinforcing spirals: The mutual influence of media selectivity and media effects and their impact on individual behavior and social identity. *Communication theory* 17, 3 (2007), 281–303.
[14] Haiyi Zhu and Bernardo A Huberman. 2014. To switch or not to switch: understanding social influence in online choices. *American Behavioral Scientist* 58, 10 (2014), 1329–1344.
[15] Cai-Nicolas Ziegler, Sean M McNee, Joseph A Konstan, and Georg Lausen. 2005. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web.* ACM, 22–32.