User Preference Dynamics on Evolving Social Networks - Learning, Modeling and Prediction

Fabiola S. F. Pereira
Federal University of Uberlândia
Uberlândia, Brazil
fabiola.pereira@ufu.br

Gina M. B. Oliveira
Federal University of Uberlandia
Uberlandia, Brazil
gina@ufu.br

João Gama
INESC TEC, University of Porto
Porto, Portugal
jgama@fep.up.pt

ABSTRACT

The preferences adopted by individuals are constantly modified as these are driven by new experiences, natural life evolution and, mainly, influence from friends. Studying these temporal dynamics of user preferences has become increasingly important for personalization tasks. Online social networks contain rich information about social interactions and relations, becoming essential source of knowledge for the understanding of user preferences evolution. In this thesis, we investigate the interplay between user preferences and social networks over time. We use temporal networks to analyze the evolution of social relationships and propose strategies to detect changes in the network structure based on node centrality. Our findings show that we can predict user preference changes by just observing how her social network structure evolves over time.

KEYWORDS

temporal networks, social network analysis, user preferences, evolutionary network analysis

1 INTRODUCTION

This paper is an extended abstract of the achieved results of the thesis defended in the Postgraduate Program in Computer Science of the Federal University of Uberlândia in July 18, 2018. The work was developed by the first author, during 53 months, under supervision of the last authors.

What drives people's preferences dynamics? Modeling users' preferences and needs is one of the most important personalization tasks in information retrieval and recommender systems domains. User preferences are fairly dynamic, since users tend to exploit a wide range of items and modify their tastes accordingly over time. Moreover, all the time users are facing with others' opinions and suffering social influence.

Given the above context, the main purpose of the thesis is to investigate *the interplay between user preferences and social networks over time for systems personalization, through the online prediction of user preference change events on a continuously evolving network*. We hypothesize that the evolution of a user's preference is related to the evolution of his/her social network structure, specially when it comes to the detection of changes.

As motivation example, let us consider a context concerning news that users like to read in everyday life. Suppose that analyzing the preferences of a given user A, we detect that on Aug

21st, A prefers to read about *politics* and *economy* than other news categories such as *sports* or *health*. Then, in a second moment, A's preferences remain stable, just appearing a preference of *politics* over *economy* news. However, in a third moment, on Aug 30th, we observe that A's preferences have changed and now, *economy* is preferred over *politics*. This situation is illustrated in the upper part of Figure 1, where preferences are represented by better-than graphs (a directed edge (u, v) indicates that u is preferred over v). In the lower part of Figure 1, snapshots of A's social network are represented. We notice that the network is also evolving with nodes appearing, disappearing, associating and disassociating with each other as time flies. In the network, nodes are Twitter users and a directed edge (x, y) means that y retweeted[1] x, i.e., the information flow. We conjecture that many aspects of A's social network can influence on A's preferences evolution. For instance: around Aug 27th, A was being influenced by users who also like *politics*. From 27th to Aug 30th, a new connection with an influential personality in *economy* may have appeared and influenced A. A is always in contact with people who like *sports*.

It is an essential point to detect and predict A's preferences evolution and changes over time. We show in this thesis that the temporal-topological social network structure of a given user is strongly correlated with her preference dynamics. According to our findings, by just observing A's social network evolution, we could increase the assertiveness of a news recommendation system for example, when recommending economy instead of politics news to A from Aug 30th.

2 RESEARCH CHALLENGE

Our work is related to a number of areas of study, including (i) preference dynamics, (ii) user preferences in social networks and (iii) event detection in evolving networks. The research challenge lies at the junction of these topics. In this chapter we identify, organize and discuss the state of art. Figure 2 is an illustrative schema showing how we position in related research. In literature these topics are combined two by two what leads to the research challenge of our proposal by combining all of them.

Most of the works exploring user preferences evolution focus on time-aware recommendation considering user profiles variation over time. There are also approaches performing analytical evolution to describe user preferences, behavior, sentiments and actions [8]. None of them, however, considers social networks. Works combining preferences with social networks seek to discover how these preferences are extracted from the network [1]. Related work here varies a lot according to data domain, but does not explore the time

---

[1]Retweet is to share some content originally posted by someone else in Twitter.
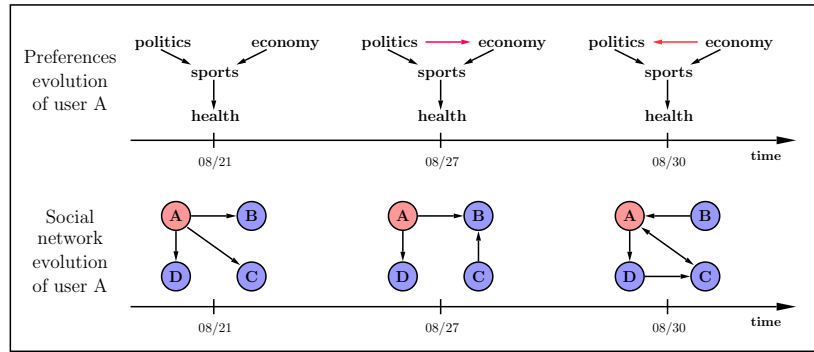
Figure 1: Evolving perspective of A's temporal preferences (top) and A's temporal social network (bottom). Preferences are represented by better-than graphs where an edge (u, v) means that u is preferred over v. In the network, nodes are Twitter users and an edge (a, b) means that b retweeted a. On 08/30 there was a preference change. At the same time, A significantly changed her structural position (node centrality) on the network.
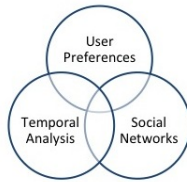
Figure 2: Main related research topics of this thesis. The originality lies at the junction of these three topics.

dimension. Finally, a large collection of works search for events in social streams using techniques of text mining. Another direction, aligned with our proposal, detects events in graphs that evolve their structure [2, 14, 24]. The challenge under this perspective is to apply those techniques to the user preference dynamics problem.

3 APPROACH AND UNIQUENESS

User Preference Dynamics (UPD) refer to the observation of how a user evolves his/her preferences over time. This evolution can be related with users becoming more or less demanding, with concept drifts in their preferences or even with some anomalous tastes at certain times. The hypothesis of this work is founded on the idea that social networks are good source of knowledge to investigate how user preferences evolve over time. In what follows we delineate our methodology.

Temporal Dynamics of User Preferences. The first step was to define what are user preference dynamics (UPD). We have defined a temporal preference model able to describe user preferences over time through user profiles. Then, we have proposed an algorithm to detect changes on temporal preferences as time evolves. The *PrefChangeDetection* algorithm is founded on the analysis of *better-than graphs* (BTG), which are graphs composed by preference domain objects of a given user, during some observation period T (for full description, please refer to Algorithm 1 in [16]). From proposed definitions and having in mind our goal, we have stated our main problem as:

given (i) a user u, (ii) an evolving social network $\mathcal{N}$ and (iii) the set of objects in preference domain A, predict preference changes events $\delta_{t'}^{u}$ over A, for any $t' > t$.

Temporal Networks for UPD analysis [17]. As problem solution, we have begun investigating temporal networks [9, 10]. We have proposed modeling a follower/followee Twitter network as a temporal social network. In this analysis the goal was to perceive how nodes evolve as a function of temporal centrality metrics. We have revisited the concept of shortest path considering the time dimension [20]. We have shown how to compute closeness and betweenness centralities using fastest paths. We have developed an algorithm for mining streams of temporal networks and have used it to perform analysis over Twitter follower/followee network. Our findings have shown that analyzing Twitter as a temporal graph, models the behavior of real applications and is different from just considering static analysis.

Node Event Detection in Evolving Social Networks. After temporal networks and evolving centralities in Twitter analysis, we have moved one more step towards problem solution. We have explored the idea of centrality-based node event detection on evolving networks [2, 3, 5, 21]. The goal was to detect at what points in time a node change its behavior significantly. Our proposal was a node event mining model with two different strategies for detecting change points – based on *average score* of nodes' centrality values inside a time window W and based on *ranking score* of nodes, which is the acceleration of a node in the ranking position from the past to current instant time, regarding node centrality value.

Finally, for our node event mining task we propose a sliding window based algorithm that is able to detect node events as the network evolves (Algorithm 3 in [16]). The most common way of processing evolving networks is assuming them as edges streams. The idea is to work with a sliding window strategy based on time instants. Thus, as time goes by, the oldest edge stream objects are forgotten. According to our model, the node event detection is based on centralities functions. Data summarization strategy varies according to the change-point scoring function choice.

We have also performed an empirical analysis using a Twitter interaction network (retweets), with the intention to observe events detected with closeness centrality and perceive the relation of detected events with network semantics (Figure 3).

Correlating changes on user preferences and changes on node centrality in temporal networks [15, 18]. We then joined our findings and proposals and performed an experimental evaluation focused on our main goal: the interplay between user preferences and social networks over time.

Table 1 gives some basic statistics of the networks we used for studying UPD and node events: Twitter and Jam networks. We built a Twitter interaction network, where nodes are Twitter users and an edge (e_1, e_2, t) represents that e_2 retweeted at t some text originally posted by e_1. As network links are retweets, we assume that contacts are instantaneous considering a minimum time granularity of 1 day. In order to discover what users are talking about on the network we performed topic modeling with the LDA algorithm [7]. $A = \{politics, international, corruption, sports, security, education, en-tertainment, economy, religion, others\}$ is the set of objects in the domain on which we extract user preferences and each tweet is labeled with one object $o \in A$

This Is My Jam (TIMJ) was an online social music network where users could share their favorite songs with their followers. Only one song could be shared at a time – the current jam, which lasted for up to one week in users' status. Furthermore, as a social network, users could like each other's jam. TIMJ dataset was released by [11]. We built a temporal network based on users' likes, where nodes are Jam users and an edge (u_1, u_2, t) means that u_2 liked u_1's jam posted at t. In this way, the directed edges represent the music influence flow.

Table 1: Summary of networks statistics

	Twitter network	Jam network
Domain	Brazilian news in Twitter	social music network
Time span	08/08/2016 - 11/09/2016	08/26/2011 -09/26/2015
# nodes	292,310	54,393
# temporal edges	1,392,841 (retweets)	1,667,335 (likes)
Avg static path length	12.31	7.63
Avg temporal path length	5 (day granularity)	2 (week granularity)

We have discovered that there is a correlation between preference change events and centrality-based node events, specially when considering temporal networks. Moreover, we have concluded that temporal closeness centrality is more suitable when correlating UPD and node events than betweenness. Figure 4 illustrates a sample of our results highlighting the comparison between static and temporal networks correlation strengths.

The Preference Change Prediction Model [19]. Lastly, we have developed a complete solution for the preference change prediction problem, taking into account efficient strategies to detect node events in evolving networks. The model processes evolving network streams and detect change points based on centrality measures. We have explored two window-based aggregating mechanisms – moving window average (MWA) and weighted moving window average (WMWA), and a third memory less mechanism, the Page-Hinkley (PH). Moreover, we have implemented algorithms considering degree, betweenness and closeness centrality measures. We have

applied our proposed model in the user preference change detection problem and evaluated the performance of our algorithms on homogeneous and bipartite Twitter networks. As result, we got promising predictive performances, reaching F1=0.85.

Remarking on Figure 5 we summarize our proposed schema for UPD analysis with evolving social networks focusing on preference change prediction.

4 RESULTS AND CONTRIBUTIONS

Our research produced a series of original contributions to the fields of social network analysis, dynamic network analysis and preference learning. We highlight:

- introduction and formalization of the UPD problem focusing on preference changes;
- an innovative analysis of users' temporal centralities evolution in Twitter network modeled as a temporal social network;
- a new approach for centrality-based events detection in continuous evolving networks;
- discovery of a correlation between user preference changes and centrality-based events in temporal networks;
- a preference change prediction model able to predict preference changes events from a continuously evolving social network.

We made available all datasets and softwares built and used in this thesis[2]. Table 2 summarizes our datasets. Those related with Twitter were first collected from crawlers. The This Is My Jam dataset was released by [11] and we made available the resultant temporal network. Source codes refer to (i) Twitter crawlers, (ii) a Gephi [6] temporal centrality plugin, (iii) a node event detection algorithm and (iv) user preference extractors.

Table 2: Summary of networks produced in this thesis. ITV: interval graph; CT: contact sequence graph; BP: bipartite.

| Dataset	Definition	Type	Chapter
Twitter celebrities	followers	ITV	5
Twitter Brazilian News	retweets	CT	6
This Is My Jam	likes	CT	7
Twitter Brazilian News	retweets	CT	7, 8
Twitter Brazilian News BP	retweets/ topics	CT	8

5 DISCUSSION

The work presents relevant contribution to *social network analysis*, *web user preferences* and *temporal networks processing*. We proposed a temporal preference model able to describe user preferences over time and to detect changes on temporal preferences. By correlating changes on preferences and changes on node centralities we move towards understanding how *content* and *topology* evolve on a social network. Our proposed predictive model is able to detect preference changes by just observing a continuously evolving network.

The new methodologies described in this thesis can be improved and extended in a number of ways, opening several opportunities for future research. We outline some ideas: (i) development of evaluation metrics without ground truth in social networks [22], (ii)

---

[2]https://github.com/fabsfernandes/upd-evolvingnetworks
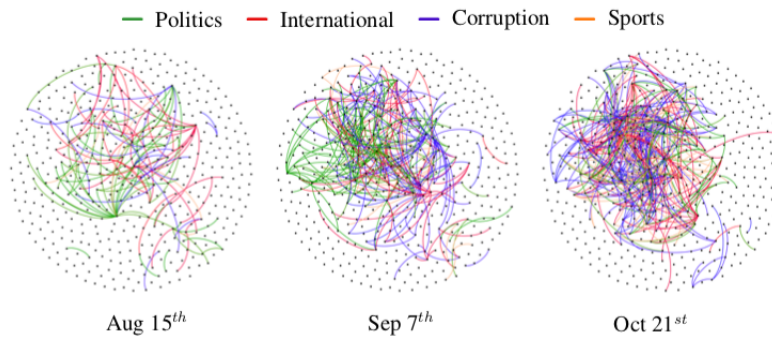
Figure 3: Snapshots of samples of the evolving interaction network (Aug 15th, Sep 7th, Oct 21st). Nodes are Twitter users. One tie from user u_1 to u_2 means that u_2 retweed at t some text originally posted by u_1. The colors represent topics that users are talking about at t. Samples were built by filtering nodes with degree between 50-22000 and edges representing the 4 most popular topics. Each snapshot corresponds to 1 day time-interval. This figure highlights the *edges* evolving aspect. Nodes are not evolving for better visualization.
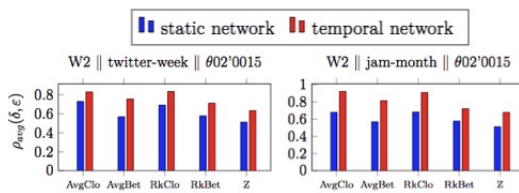
Figure 4: *Pearson* correlation coeficient between preference changes δ and centrality-based node events ϵ for (left) Twitter dataset and (right) TIMJ dataset. AvgClo: average score by closeness; AvgBet: average score by betweenness; Rk-Clo: ranking score by closeness; RkBet: ranking score by betweenness; Z score [4]; $W2$: window size of 2 time units; θ: threshold for events.
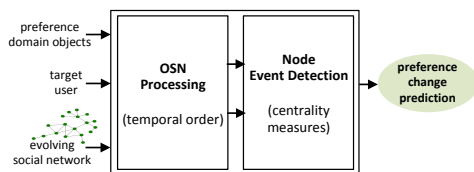
Figure 5: Proposed schema for UPD analysis with evolving social networks focusing on preference change prediction. OSN: online social network.

techniques for stream processing centrality metrics in networks [13] and (iii) models for descriptive preference change prediction [12, 23].

REFERENCES

[1] Mohammad Ali Abbasi, Jiliang Tang, and Huan Liu. 2014. Scalable Learning of Users' Preferences Using Networked Data. In *25th ACM Conf. Hypertext and Social Media (HT '14)*. 4–12. https://doi.org/10.1145/2631775.2631796
[2] Charu Aggarwal and Karthik Subbian. 2014. Evolutionary network analysis: a survey. *Comput. Surveys* 47, 1 (2014), 10–36.
[3] Charu C. Aggarwal and Karthik Subbian. 2012. Event Detection in Social Streams. In *12th SIAM International Conference on Data Mining, USA*. 624–635.
[4] Leman Akoglu and Christos Faloutsos. 2010. Event detection in time series of mobile communication graphs. In *Army Science Conference (18)*, Vol. 2. 1–8.
[5] Leman Akoglu, Hanghang Tong, and Danai Koutra. 2015. Graph based anomaly detection and description: a survey. *Data Mining and Knowledge Discovery* 29, 3 (2015), 626–688.
[6] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. 2009. Gephi: An Open Source Software for Exploring and Manipulating Networks. *AAAI Conference on Weblogs and Social Media* (2009).
[7] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* 3 (March 2003), 993–1022.
[8] Anaïs Cadilhac, Nicholas Asher, Alex Lascarides, and Farah Benamara. 2015. Preference Change. *J. of Logic, Language and Information* 24, 3 (2015), 267–288.
[9] P. Holme. 2014. Analyzing Temporal Networks in Social Media. *Proc. IEEE* 102, 12 (2014), 1922–1933.
[10] Petter Holme and Jari Saramaki. 2012. Temporal networks. *Physics Reports* 519, 3 (2012), 97–125.
[11] Andreas Jansson, Colin Raffel, and Tillman Weyde. 2015. This is my Jam – Data Dump. *16th International Society for Music Information Retrieval* (2015), 1–2.
[12] Komal Kapoor. 2014. *Models of dynamic user preferences and their applications to recommendation and retention.* Ph.D. Dissertation. University of Minnesota.
[13] Miray Kas, Matthew Wachs, Kathleen M. Carley, and L. Richard Carley. 2013. Incremental Algorithm for Updating Betweenness Centrality in Dynamically Growing Networks *(ASONAM)*. 33–40.
[14] Gabriel P. Oliveira, Natércia A. Batista, Michele A. Brandão, and Mirella M. Moro. 2018. Tie Strength in GitHub Heterogeneous Networks. (2018), 363–370.
[15] F. Pereira, J. Gama, S. Amo, and G. Oliveira. 2018. On Analyzing User Preference Dynamics with Temporal Social Networks. *Mach. Learning* (2018), 408–423.
[16] Fabiola S. F. Pereira. 2018. *User Preference Dynamics on Evolving Social Networks – Learning, Modeling and Prediction.* Ph.D. Dissertation. UFU.
[17] Fabiola S. F Pereira, Sandra Amo, and Joao Gama. 2016. Evolving Centralities in Temporal Graphs: a Twitter Network Analysis. In *Mobile Data Management (MDM), 2016 17th IEEE International Conference on.* 43–48.
[18] Fabiola S. F. Pereira, Sandra de Amo, and Joao Gama. 2016. On Using Temporal Networks to Analyze User Preferences Dynamics. In *Discovery Science: 19th International Conference, DS 2016, Bari, Italy, 2016.*
[19] Fabiola S. F. Pereira, Shazia Tabassum, and Joao Gama. 2019. *Processing Evolving Social Networks for Change Detection Based on Centrality Measures.* Springer International Publishing. 155–176 pages.
[20] Huanhuan Wu, James Cheng, Silu Huang, Yiping Ke, Yi Lu, and Yanyan Xu. 2014. Path Problems in Temporal Graphs. *VLDB Endowment* 7, 9 (2014), 721–732.
[21] Reza Zafarani, Mohammad Ali Abbasi, and Huan Liu. 2014. *Social Media Mining: An Introduction.* Cambridge University Press, New York, NY, USA.
[22] Reza Zafarani and Huan Liu. 2015. Evaluation Without Ground Truth in Social Media Research. *Commun. ACM* 58, 6 (2015), 54–60.
[23] Jun Zhang, Chaokun Wang, Jianmin Wang, and Jeffrey Xu Yu. 2014. Inferring Continuous Dynamic Social Influence and Personal Preference for Temporal Behavior Prediction. *Proc. VLDB Endow.* 8, 3 (2014), 269–280.
[24] M. Zignani, S. Gaito, and G. P. Rossi. 2018. Follow the Mastodon: Structure and Evolution of a Decentralized Online Social Network. In *ICWSM*. 541–550.