Twixonomy Visualization Interface: How to Wander Around User Preferences

Giorgia Di Tommaso, Giovanni Stilo

Sapienza University
Rome, Italy
ditommaso@di.uniroma1.it, stilo@di.uniroma1.it

Abstract. User interfaces have become essential tools for a user to interact with a recommender system. In the two most common settings, the user interface either helps users to collect their preferences, or to provide an explanation of the generated recommendations. In this paper, we present the *Twixonomy Visualization Interface*, a tool that allows users both to explore their preferences and to discover new ones. Preferences are represented by a Wikipedia Category DAG connected with the initial (primitive) preferences implicitly or explicitly expressed by the user. Our tool can be considered as an integration of a recommender system since, by exploring the DAG, the user can both analyse the connections between his/her preferred items and other semantically related items or categories, and understand the motivations for new, serendipitous recommendations.

Keywords: Recommender system; user interface; twitter; social media; data mining; graph; pruning; semantic; user profiling.

1 Introduction

The scope of Recommender System (RS) is to provide suggestions, or directing a person to a service, product or content of interest [3]. It is widely-known that the effectiveness of a recommender systems cannot only be measured by its accuracy, but there are other elements that play a role, in order for the user to trust the system and perceive it as an advice-giver [22]. In his ACM "Intelligent User Interfaces" tutorial [10], Konstan highlighted the relevance that the user interface has, in order to make a recommender system understandable by the users. Indeed, when a user can visualize and contextualize the recommendations with her preferences, it is more likely that she will accept what is recommended to her.

In [4, 5], we presented a method for large-scale analysis of Twitter users supported by a hierarchical representation of their interests, which we call a Twixonomy. In order to build a population, community, or single-user Twixonomy we first associate "topical" friends in users' friendship lists (i.e. friends representing an interest rather than a social relation between peers) with Wikipedia categories. A word-sense disambiguation algorithm is used to select the appropriate

wikipage for each topical friend. Starting from the set of wikipages representing the main topics of interests of the initial Twitter population, we extract all paths connecting these pages with topmost Wikipedia category nodes, and we then prune the resulting graph efficiently so as to induce a direct acyclic DAG graph and significantly reduce over ambiguity. This graph is the Twixonomy.

In this paper, we are going to present a tool, called Twixonomy Visualization Interface (TVI), which allows a user to explore his/her Twixonomy, understanding what his/her interests are, what are the high-level Wikipedia categories associated to these preferences, and the relationship between these preferences. Moreover, it allows to compare the Twixonomy of a user with that of another Twitter profile, thus having also a list of the shared interests in terms of categories, and a compact similarity measure (known as *affinity score*) between the two Twixonomies.

By allowing the users to visually analyze their preferences and discover new ones in terms of categories they are associated to will be useful from multiple perspectives. Indeed, by seeing the correlation between the categories represented in the Twixonomy and the preferences they explicitly expressed, it is more likely that the user will trust the system and accept the categories as recommendations of something they might be interested in. As a consequence, these categories can be directly employed as an information source by a recommendation algorithm, thus broadening the preferences of the users and generating more rich recommendations, certainly involving new items and possibly also surprising (serendipitous) ones.

2 Related Work

This section presents the works related to our purpose, mostly focusing on recommender system in social networks and on the role of the user interface when producing recommendations to the users.

Recommender system in social networks In the Social Networks literature there is a considerable number of works in which users' interests are extracted for some specific purpose, like detecting trending topics, i.e., topics that emerge and become popular in a specific time slot. Trending topics are extracted to produce a recommendation [11, 6, 13], to model users' expertise [21], or to detect shared interests (e.g., events) that are predominant in a given time span [12]. The large majority of scholars are concerned with user recommendation, a task for which many solutions have been presented in literature (see [11] for a survey); since they model user interests regardless of specific applications.

The studies more closely related to our work are [8] and [2]. In the former, the authors aim to derive a categorial representation of users' interests using Wikipedia. First, named entities are extracted from the text in micro-blogs, then, Wikipedia pages, named *primitive interests*, are associated to each named entity. To select a reduced number categories, denoted as *hierarchical interests*, spreading of activation [1] is used on a pruned version of the Wikipedia graph,

where active nodes are initially the set of primitive interests. Note that, despite their name, hierarchical interests are not hierarchically ordered: rather, they are a set of mid-high level categories. The second related work is [2], where, similarly to us, users' interests are inferred at a large scale. This system, named Who Likes What, was the first system capable of inferring Twitter users' interests at the scale of millions of users. First, the topical expertise of popular Twitter users is learned from the names and descriptions of Twitter lists in which such users actively participate. Then, the interests of users who subscribe to at least 3 expert users are transitively inferred. Who Likes What is also deployed online with a web interface; we are going to analyze further which are the differences and advances between the two systems.

User interface and recommender systems The user interface plays a crucial role in the way the users perceive the recommendations [16]. Indeed, it is either employed in a recommender system to collect preferences from the users [15], or to provide a better understanding of the recommendations and allow a user to explore them [18, 20] (our proposal fits with the second scenario). In [19], a user-centric framework, called *ResQue* measures the user experience with a recommender system. Knijnenburg et al. [9] studied five interaction methods of the users with a recommender systems and showed that users have different preferred methods of interaction. *SmallWorlds* [7] provides a graph-based explanation of the recommendations generated with a collaborative filtering approach. *SetFusion* [17] gives the users the possibility to control a hybrid recommender system, by setting the relevance that each component should have when generating the recommendations.

Contributions Our approach provides a user interface to allow the user to get in touch and explore his/her preferences. An interface that provides this useful informations was never been proposed. With respect to the analyzed bibliography, our main contribution is the visualization and the comparison between one or more Twixonomies, providing the following advantages:

- it allows to visualize the interests of single users, communities and populations in a easily browsable and interpretable way, especially when compared with the large number of unstructured and fine-grained topics extracted from textual features in messages and lists descriptions (like in [2]);
- it allows to compare several Twixonomies giving a precise measure of semantic symilarity;
- it is also designed as a predictive system that aims to discover the interests not declared directly from the users and to recommend new ones on these basis, also in the presentation aspect.

3   Proposed Approach

This Section first briefly describes the Twixonomy algorithm as described in [4], and then discuss the functionalities exposed by Twixonomy Visualization Interface (TVI).

3.1   Twixonomy Algorithm

Figure 1 provides a general picture of what the Twixonomy Algorithm produces. For every user p (@JohnDoe) of a population P with a Twitter account, we consider his followees (@OrlandoMagic, @Newsweek, @JohnMom, etc.). Some of these followees are genuine friendship or parental relations (@JohnSister, @JohnMom), some are possibly non-reciprocated relations directed towards "authoritative" Twitter accounts, indicating an interest of our user, rather than a peer relation. In this case, it is likely that such followees have a correspondent Wikipedia page, as shown in Figure 1. Starting from these pages, we use the Wikipedia Category Graph to induce a full-fledge taxonomy, the @JohnDoe's interests Twixonomy. In Figure 1, dotted edges represent longer hypernymy paths in the Twixonomy, that have been cut for the purpose of space.

A Twixonomy can be created both for single users, communities, and for the entire Twitter population. Since a category in a Twixonomy can be associated to all the users sharing that interest, the Twixonomy enables semantic group profiling and recommendation.

Given a population of Twitter users P (possibly, $|P| = 1$), we generate the Twixonomy in three steps:

1. First, we extract from the profiles of each user p in P the set F of followees, such that each u in F is followed by at least one user p in P. Note that the sets P and F are different, though possibly overlapping. We then create a mapping between users $u \in F$ and wikipages in Wikipedia, if any;
2. Next, we extract all paths connecting these pages with topmost Wikipedia category nodes in the Wikipedia Category Graph, thus inducing a graph WG;
3. Finally, we apply a graph pruning algorithm to remove cycles and reduce *multiple inheritance*, a problem that often cancels the advantage of adding semantics [14].

3.2   Functionalities

The Twixonomy Visualization Interface provides the following four main functionalities:

1. ***Search Bar*** allows to the user of the visualization interface, to search for a specific Twitter's profile. The user inserts into the search bar a query string and the system shows a list of Twitter's profiles that best match the requirements. The user can select one of the proposed profiles to visualize the related Twixonomy. The search bar is presented to the user in the landing page and it is also available in all the other views of the interface.
2. ***Graph Visualization*** represents the core of the Twixonomy visualization Interface. The Twixonomy is visualized in this part of the view (bottom-right of Figure 2) as a graph and it is always possible to select between these three sub-functionalities:
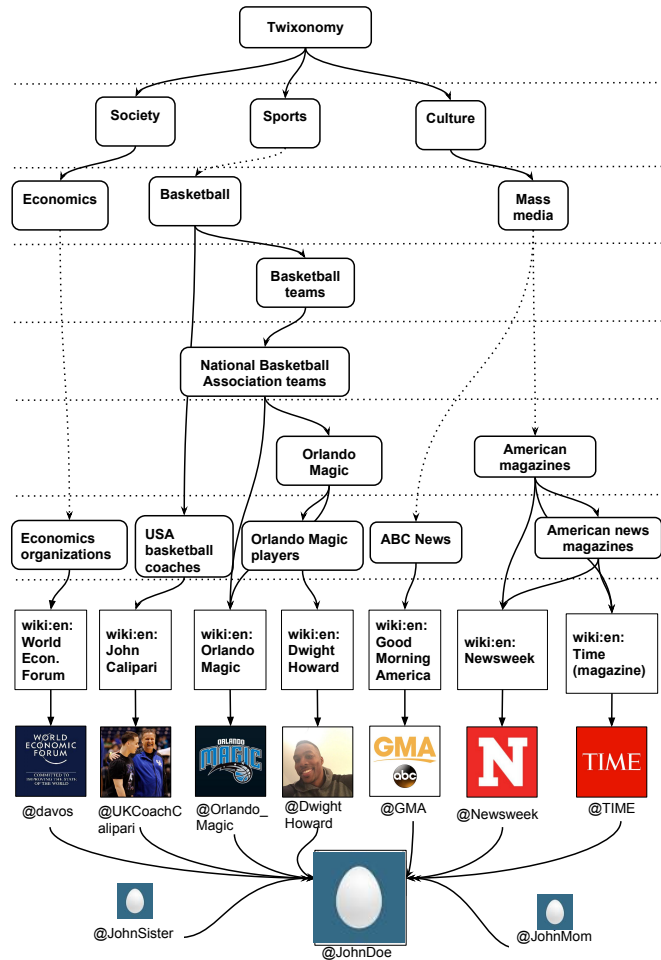
Fig. 1. Example of Twixonomy for a single user

(a) *Network* displays the Twixonomy graph of the selected user exactly as showed in Figure 2. In this type of visualization, nodes are disposed using a circular layouts. Each type of node has its own logo: nodes with Twitter logo indicate the topical friends; the associated Wikipage nodes are identify by the Wikipedia W logo, and the other nodes in the network are those inferred from Twixonomy. In our example it is possible to note that the principal interest of the selected Twitter's profile is related to the photography domain.

(b) the *Hierarchical* view allows a user to navigate the Twixonomy of the selected profile using a hierarchical layout. Nodes are presented without any icons and only the text label is associated to each node. Nodes are
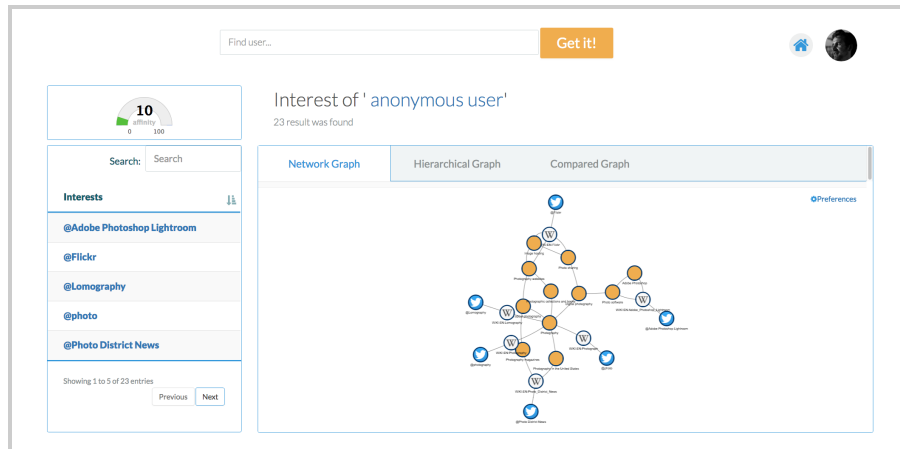
Fig. 2. Snapshot of Twixonomy Visualization Interface

organized in levels; this view is designed to make it easier to understand the relations between nodes.

(c) the *Compared* view allows a user to navigate the Twixonomy of two selected profiles at the same time. Nodes of each Twixonomy are disposed using circular layouts and the same icons of the *Network* view are used. This view it is designed to better understand what are the semantic differences and the common points between two profiles.

For all the above exposed views (*Network, Hierarchical, Compared*), it is possible to zoom in and zoom out the graph, move nodes and edges and pan the view.

In almost every case the Twixonomies are too big to allow a readable navigation of the graph. To deal with this issue we allow the user of the platform to select the depth level of the visualized Twixonomy.

For each interest node is also possible to read the related Wikipedia description and it is also possible to share the visualized Twixonomy over the principals social media platforms (Twitter, Facebook).

3. **Interest List** is presented on left part of the interface and contains all the nodes of the Twixonomy when the selected view is *Network* or *Hierarchical* . When the selected view is *Compared* only the nodes shared between the compared Twixonomies are displayed. In both cases the *Interest List* can be downloaded as a textual file.

4. **Affinity Score** is presented to the user on the top left part of the interface (see Figure 2), when two Twixonomies are selected (*Compared* functionality). The *Affinity Score* is the compact representation of the semantic similarity between the selected profiles. The semantic similarity is computed using the

following formula as proposed in [4]:

$$SemSym(A, B) = \frac{\sum_{i=1}^{N_k} w(d_{A_i^k}) \times w(d_{B_i^k})}{\sqrt{\sum_{i=1}^{N_k} w(d_{A_i^k})^2} \times \sqrt{\sum_{i=1}^{N_k} w(d_{B_i^k})^2}} \tag{1}$$

In the formula, A and B are the semantic vectors, extracted from each Twixonomy, associated to users a and b , A_i^k, $i = 1 \ldots n_k$, is the i-th boolean argument of A and is non-zero if the Twixonomy of a includes the node c_i^k of the population's Twixonomy. The index $k = 0 \ldots K$ is the generalization level ($k = 0$ indicates Wikipages, as shown in Figure 1), that we also denote as L_k, and $n_k = |V_k|$ is the total number of nodes in the Twixonomy up to L_k. Furthermore, $d_{A_i^k} = k$ is the length of the minimum path connecting c_i^k with a leaf node[1]. Finally $w(d) = \beta \cdot e^{-\alpha \cdot (d+1)}$ is a weight function with exponential decay[2], where we empirically set $\beta = 2$ and $\alpha = 0.5$. In formula (1), non-zero terms in the numerator are those for which $A_i^k = B_i^k$, however the contribution of a match exponentially decays with the distance k of matching categories from leaf nodes. The higher is the value of SemSym, the more are the shared interests between the selected users.

4   Recommendation Purpose

As we previously remarked, user interface plays a crucial role in the way the users perceive the recommendations. Indeed, it is either employed in a recommender system to collect preferences from the users, or to provide a better understanding of the recommendations and allow a user to explore them. The Twixonomy Visualization Interface is designed not only for descriptive analysis of the interest inferred from Twitter users profile but and even as a predictive system that aim to discover the interests not declared directly from the users and to recommend new ones on these basis.

    In a typical use-case, a user starts the investigation process by visualising her own Twixonomy using the *Network* view (point 2a of the previous section). With this type of visualization, a user should discover intuitively (also guided by different icons) which are the categories that dominate his interests. Using this visualisation mode is then possible, searching in the *Interest List*, to find a specific interest even if the Twixonomy is huge. By taking a closer look to a specific node, it is then possible to understand how strong is the relation with neighbor nodes. If the user would like to know which is the generality level of one node, he can switch to the *Hierarchical* view (point 2b). Indeed, by using the *Hierarchical* view, it is possible to understand what is the type of relationship of the Twixonomy and see which categories are higher than others.

---

[1] note that leaf nodes matches have $d = 0$

[2] with w(d) we mean generically either $w(d_{A_i^k})$ or $w(d_{B_i^k})$

To inspect a second use-case, suppose that a user wants to understand how close he is (in term of interests) to another Twitter's profile. Then is possible to select the *Compared* view (Secton 2c) to immediately take a look at two Twixonomies at the same time. The user then can visually inspect what are the different or the shared interests between the compared users. If it is not possible to effectively understand which are the similar interests of the Twixonomies (due to the large amount of interest of both profiles), then the *Interest List* helps the user to have a more compact representation of the similarity exposed by the profiles, by highlighting only the shared interests. Simultaneously, the *Affinity Score* offers to the user the fastest way to understand what is the level of similarity exposed by the selected profiles.

At the end of the process the user can understand not only his preferences but also how these are generated, what is the level of generalisation between them, and which of them depends by others. User that also have to compare their interests to those of another Twixonomy, can also understand which is their affinity to another profile and which are the potential new preferences that they should be interested in.

5   Conclusion and future work

In this paper, we presented the Twixonomy Visualization Interface, a tool to explore and analyze a hierarchical representation of the user preferences, called Twixonomy. Compared to [2], that presents users' preferences as simple tag-cloud, our interface presents to the user a more detailed informations; in addition Twixonomy Visualization Interface enables to wonder around the interest and compare own interests with those of the community. We are actually thinking of many possible improvements of the platform. First, we plan to embed the Twixonomy Visualization Interface in the Twitter platform in a way that should be easily possible to build the Twixonomy of a community (*list*, in Twitter terminology). Another improvement that we plan to develop is to give the possibility to each user to evaluate their Twixonomy by expressing a degree of precision and allowing feedbacks on the quality of the interface. Another on-going study regards the user interface evaluation. In particular, we would ask to users which views is more useful between the ones that are already available and the ones that are under developments.

References

1. H. Anderson. Fire spread and flame shape. *Fire Technology*, 4(1):51–58, 1968.
2. P. Bhattacharya, M. B. Zafar, N. Ganguly, S. Ghosh, and K. P. Gummadi. Inferring user interests in the twitter social network. In *Proceedings of the ACM Conference on Recommender Systems*, RecSys '14, pages 357–360, 2014.
3. R. Burke. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4):331–370, 2002.

4. S. Faralli, G. Stilo, and P. Velardi. Large scale homophily analysis in twitter using a twixonomy. In *Proc. of the 24th Int. Joint Conf. on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, pages 2334–2340, 2015.
5. S. Faralli, G. D. Tommaso, and P. Velardi. Semantic enabled recommender system for micro-blog users. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, pages 992–998, Dec 2016.
6. R. Garcia and X. Amatriain. Weighted content based methods for recommending connections in online social networks. In *The 2nd ACM Work. on Recommendation Systems and the Social Web*, Barcelona, Spain, 2010.
7. B. Gretarsson, J. O'Donovan, S. Bostandjiev, C. Hall, and T. Höllererk. Small-worlds: Visualizing social recommendations. In *Proceedings of the 12th Eurographics / IEEE - VGTC Conference on Visualization*, EuroVis'10, pages 833–842, Chichester, UK, 2010. The Eurographs Association & John Wiley & Sons, Ltd.
8. P. Kapanipathi, P. Jain, C. Venkataramani, and A. Sheth. User interests identification on twitter using a hierarchical knowledge base. In *The Semantic Web: Trends and Challenges*, volume 8465, pages 99–113. Springer Int. Publishing, 2014.
9. B. P. Knijnenburg, N. J. Reijmer, and M. C. Willemsen. 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*, RecSys '11, pages 141–148, New York, NY, USA, 2011. ACM.
10. J. A. Konstan. Tutorial / hci for recommender systems: A tutorial. In *Proceedings of the 16th International Conference on Intelligent User Interfaces*, IUI '11, pages 463–, New York, NY, USA, 2011. ACM.
11. S. Kywe, E.-P. Lim, and F. Zhu. A survey of recommender systems in twitter. In *Social Informatics*, volume 7710 of *Lecture Notes in Computer Science*, pages 420–433. Springer Berlin Heidelberg, 2012.
12. Q. Li, J. Wang, Y. P. Chen, and Z. Lin. User comments for news recommendation in forum-based social media. *Inf. Sci.*, 180(24):4929–4939, 2010.
13. C. Lu, W. Lam, and Y. Zhang. Twitter user modeling and tweets recommendation based on wikipedia concept graph. Technical Report WS-12-09, AAAI Technical Report, 2009.
14. P. Matykiewicz and W. Duch. Multiple inheritance problem in semantic spreading activation networks. In *Brain Informatics and Health - Int. Conf., BIH 2014, Warsaw, Poland, August 11-14, 2014, Proceedings*, pages 252–265, 2014.
15. S. M. McNee, S. K. Lam, J. A. Konstan, and J. Riedl. Interfaces for eliciting new user preferences in recommender systems. In *Proceedings of the 9th International Conference on User Modeling*, UM'03, pages 178–187, Berlin, Heidelberg, 2003. Springer-Verlag.
16. H. Nguyen, J. Masthoff, and P. Edwards. Persuasive effects of embodied conversational agent teams. In *Proceedings of the 12th International Conference on Human-computer Interaction: Intelligent Multimodal Interaction Environments*, HCI'07, pages 176–185, Berlin, Heidelberg, 2007. Springer-Verlag.
17. D. H. Park, H. K. Kim, I. Y. Choi, and J. K. Kim. A literature review and classification of recommender systems research. *Expert Syst. Appl.*, 39(11):10059–10072, Sept. 2012.
18. P. Pu and L. Chen. Trust-inspiring explanation interfaces for recommender systems. *Know.-Based Syst.*, 20(6):542–556, Aug. 2007.
19. P. Pu, L. Chen, and R. Hu. A user-centric evaluation framework for recommender systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems*, RecSys '11, pages 157–164, New York, NY, USA, 2011. ACM.

20. L. Qiu. *Designing social interactions with animated avatars and speech output for Product Recommendation Agents in electronic commerce*. PhD thesis, Jan 2010.

21. C. Wagner, V. Liao, P. Pirolli, L. Nelson, and M. Strohmaier. It's not in their tweets: Modeling topical expertise of twitter users. In *Privacy, Security, Risk and Trust (PASSAT), 2012 Int. Conf. on and 2012 Int. Confernece on Social Computing (SocialCom)*, pages 91–100, Sept 2012.

22. K. H. Yoo, U. Gretzel, and M. Zanker. Source factors in recommender system credibility evaluation. In F. Ricci, L. Rokach, and B. Shapira, editors, *Recommender Systems Handbook*, pages 689–714. Springer, 2015.