Co-Staying: A Social Network for Increasing the Trustworthiness of Hotel Recommendations

Catalin-Mihai Barbu
University of Duisburg-Essen
Duisburg, Germany
catalin.barbu@uni-due.de

Jürgen Ziegler
University of Duisburg-Essen
Duisburg, Germany
juergen.ziegler@uni-due.de

ABSTRACT

Recommender systems attempt to match users' preferences with items. To achieve this, they typically store and process a large amount of user profiles, item attributes, as well as an ever-increasing volume of user-generated feedback about those items. By mining user-generated data, such as reviews, a complex network consisting of users, items, and item properties can be created. Exploiting this network could allow a recommender system to identify, with greater accuracy, items that users are likely to find attractive based on the attributes mentioned in their past reviews as well as in those left by similar users. At the same time, allowing users to visualize and explore the network could lead to novel ways of interacting with recommender systems and might play a role in increasing the trustworthiness of recommendations. We report on a conceptual model for a multimode network for hotel recommendations and discuss potential interactive mechanisms that might be employed for visualizing it.

Author Keywords

Recommender systems; tourism; personalization; trust; multimode networks; trustworthiness

ACM Classification Keywords

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*evaluation/methodology, graphical user interfaces (GUI), user-centered.*

INTRODUCTION

Recommender systems (RS) typically build and maintain network representations of the data they store about their users and product catalogs. One of the earliest and most well-known examples is the "item-to-item" model introduced by Amazon in its RS, which enabled the company to show, using collaborative filtering techniques, what *other* products were purchased by users who bought a certain product [8]. Linking user preferences with items and subsequently modeling the various relationships that arise between them can increase the accuracy of recommendations. Most commonly, RS employ 1-mode networks, meaning that all nodes are of the same type (e.g., users). Less frequent is the use of 2-mode networks [2], in which relationships are shaped between two different types of nodes, e.g., users and items. A relatively unexplored area of research concerns the usage of multimode (or n-mode) networks, in which vertices can be of three or even more types. To offer an example from the tourism domain, consider a hotel recommender that models the relationships that appear between users, hotels, and hotel amenities. Furthermore, the internal representation of such networks in usually not visible to users. This can happen for various reasons, ranging from privacy concerns to the inherent difficulty of creating a meaningful illustration of the structure of complex networks.

The purpose of this short paper is to advance the state of the art in two ways. First, we propose a concept of a multimode network for representing users, hotels, and hotel properties. We argue that a hotel recommender can exploit multimode networks to generate more suitable suggestions. Second, we examine potential interactive mechanisms that could be developed to facilitate the presentation of the network to users. We believe that having access to additional means of visualizing hotel information would allow users to interact in novel ways with the RS. Furthermore, such novel interaction techniques could play a role in increasing the trustworthiness of recommendations.

In the following sections, we review related work on the usage of multimode networks in RS. We also investigate some of the typical visualization techniques that have been developed so far. Next, we present our conceptual model for a user-item-attribute network using references and examples from the hotel booking domain. Finally, we discuss how allowing users to interact with such a network might introduce novel ways of interacting with RS. Additionally, we open the discussion on whether exploiting this network could play a role in increasing the trustworthiness of recommendations.

RELATED WORK

The emergence of web communities and the sustained growth of the user-generated content available online provides important opportunities for RS. Incorporating social network information has the potential to alleviate cold-start and sparsity problems, which are inherent in typical collaborative filtering approaches [18]. In the field of RS for tourism (and, more generally, for commerce), a major aspect of exploiting user-generated content is the extraction of topics and sentiments from reviews, e.g., to create richer user models [12]. Combining social networks analysis techniques, such as community detection and visualizing techniques, with RS is, therefore, a worthwhile direction for continued research [10]. A survey of current state-of-the-art

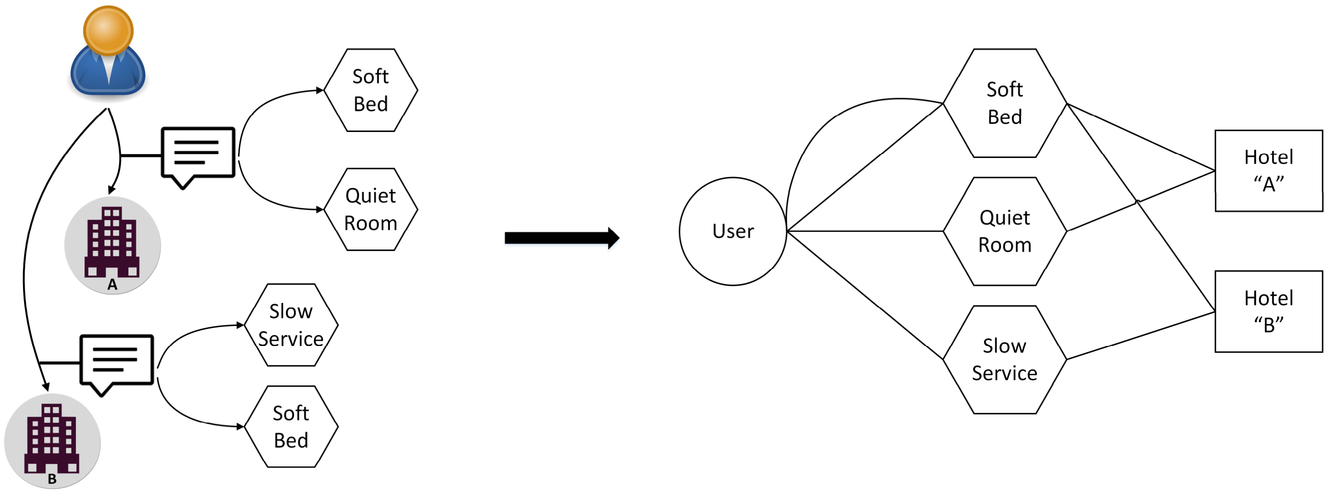

Figure 1 – Example of how a simple multimode network (right) containing three types of vertices—users (circle), topics (hexagon), and hotels (rectangle)—could be generated from crawled user-generate content (left). In our conceptual model, users and hotels are connected through topics, instead of directly.

methods in both fields as well as existing challenges are highlighted, for example, in [18].

Exploiting 2-mode networks for improving the quality of recommendations has already been achieved in various domains. One such approach was used to derive similarity information between artists and songs by exploiting data collected from a music file-sharing network [14]. The information was then used to build a 2-mode graph of users and songs, which was subsequently used to recommend new artists. Citation networks are also a frequent focus of research [9]. The usage of multimode networks for recommendations is less frequent [17]. Linking users to items via tags to improve recommendations is of definite interest in the RS community [15]. Still, tripartite graphs containing users, items, and tags have been studied more in the context of information retrieval [3]. Alternate approaches, for example using multiple 1-mode networks (as opposed to multimode networks) for generating article recommendations, have also been proposed [20].

The problem of visualizing complex networks employed in RS is also receiving increasing attention. Previous work has been done, for instance, on facilitating the exploration of recommended items by combining different entities (users, tags, and agents) [16]. Although there is ongoing research on visualization techniques for tripartite networks [7], it appears that such methods are usually not applied in a systematic way in the context of RS. We argue that multimode networks should not be used solely for enhancing the generation of recommendations. Providing means to visualize the network as well as mechanisms for interacting with it could increase the transparency and control of the RS [16]. The lack of transparency exhibited by many modern RS is frequently cited as having a detrimental effect on the perceived trustworthiness of such systems [6]. Trust is especially relevant in the context of hotel recommending, where the risk associated with poor choices is often higher. Visualizing the underlying factors used to generate recommendations could, for example, allow developers to embed *trust cues* (i.e. interface elements that allow the user to determine the reliability of the presented information) into the presentation layer [13].

CO-STAYING NETWORK

Based on our review of the literature, there appears to be a research gap with respect to the usage of multimode networks for RS. At the same time, the goal of providing interactive mechanisms for allowing users to explore the underlying graph has been studied less frequently.

Example Domain and Dataset

Hotel booking is a domain in which the trustworthiness of recommendations is especially important. More generally, while selecting the domain we considered three aspects: 1) The choice should carry a substantial amount of risk for the user; 2) the items should have a reasonable set of attributes that need to be considered; and 3) there should be a large body of user-generated content available, in the form of reviews, photos, tags, and ratings, that can be leveraged in the presentation. Because of the first criterion, we decided against using the more common domain of movie recommendations. Hotel booking, on the other hand, fulfils all three conditions.

We crawled metadata and overall 838,780 user reviews for 11,544 hotels located in five major European cities from Booking.com[1]. This real-world dataset ensures access to a representative and interesting subset of hotels, covering as many types of amenities as possible. Furthermore, it features a diverse set of reviews contributed by various types of travelers—and who are travelling for different purposes—

[1] http://www.booking.com/

thereby maximizing the variety of user opinions and arguments. A characteristic of Booking.com is that all reviews on its site are *verified*, meaning they are written by people who have stayed in those hotels. This feature reduces the number of fake reviews.

Conceptual Model

Our conceptual model, to which we will refer in the following sections as a "co-staying network", has three types of vertices: users, topics (i.e. hotel attributes), and hotels. In contrast to a typical 2-mode network, users are not linked directly to the hotels in which they stayed. Instead, an edge is first created between the user and a topic. A second edge then links the topic to the hotel for which the review was submitted (Figure 1). This process is repeated iteratively for each topic in a review as well as for all user-contributed reviews. If a review contains references to more than one topic (as is often the case in hotel reviews), there will be several paths between its author and the hotel, each passing through a topic. Furthermore, if a user mentioned the same topic in more than one review, a separate pair of edges will be created for each instance.

The crawled user reviews in the dataset are mined to extract topics. First, we identify attribute-value pairs such as "comfortable bed" in the reviews. Then, these pairs are merged with others that have the same meaning, e.g., "comfy bed". Next, using sentiment analysis, pairs are classified as positive or negative hotel properties. Finally, pairs that describe properties related to, for example, a hotel room, are classified and clustered together. Broadly, topics are divided into three main categories: room attributes (e.g., bed, shower, minibar), hotel attributes (e.g., location, swimming pool, parking), and hotel services (e.g., breakfast, Wi-Fi quality, friendliness of staff). The complete procedure is described in detail in [4]. Topic disambiguation techniques are employed to prevent duplication and correct spelling errors. For example, the terms "wi-fi", "wifi", "wireless network", "wireless internet", or "wifus" (an incorrect spelling that is encountered relatively often in user reviews) should all be filed under a single attribute, e.g., "Wi-Fi".

The user's sentiment regarding each topic is classified with respect to its *polarity* (i.e. positive, neutral, or negative) and *strength*. These two attributes are normalized and encoded in the edges as a single value in the interval [-1,1]. Values closer to the left side of the interval are indicative of strong negative sentiments about a topic (e.g., "terrible breakfast"). Similarly, higher positive values are associated with topics that a user has praised in a review (e.g., "wonderful location").

If a hotel is part of a chain, our model also allows the creation of additional "soft links" from a topic to the rest of the hotels in that chain. This is based on the premise that these hotels share many characteristics—though not all. Thus, it is likely (and typically advertised as such by hotel brands) that travelers would have comparable experiences and access to similar amenities in every location that is part of a franchise.

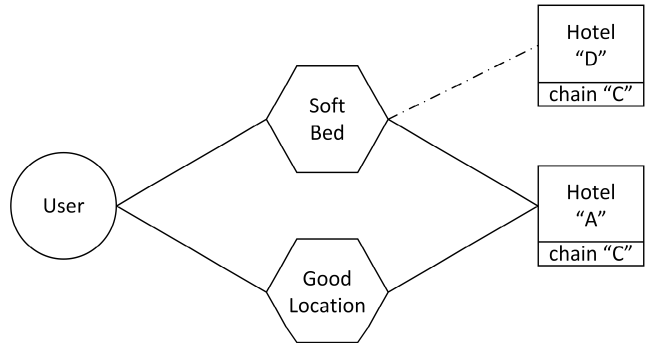

Figure 2 – Example of a soft link from a topic to a hotel that is part of the same chain as the hotel for which the review was written. Soft links can only be created for specific topics.

This approach lets the RS exploit user opinions about selected topics when suggesting recommendations that might otherwise not have enough reviews. Among the topics that could be shared in this way are those concerning room layout and furnishings, breakfast, and hotel facilities. Topics such as those related to the hotel's location or the service quality, on the other hand, would not be shared between hotels (Figure 2).

Clusters can be identified for each type of vertex in the co-staying network. Topics can be characterized by their overarching category (i.e. hotel, room, or service). Furthermore, by analyzing travel and review patterns, additional relationships between hotel topics could be identified. Hotels can be clustered based on whether they are part of a chain as well as by looking at the most common topics that characterize them. Finally, user similarity measures can be extended to include, in addition to demographic data (e.g., age range, country of origin) and rating behavior, also *experience* (e.g., based on the number of contributed reviews, frequency of travel, and the types of hotels visited). Preliminary work on these aspects has already been reported in [1].

The proposed model aims to improve upon traditional approaches to hotel recommending by creating stronger links between users, hotels, and hotel attributes. This would allow users to explore, in addition to the details of the recommended items, the public profiles and preferences of those guests whose reviews were exploited for generating the recommendations. Various forms of interactive mechanisms could be developed to facilitate this kind of exploration.

Interactive Mechanisms

The initial focal point of the network would be the hotel that is being recommended. Users might ask themselves, "What do people talk about when reviewing this hotel?" The most common topics (ideally tailored to fit the user's interests) could be shown radiating from the hotel. These should be clustered by category. Edges would connect these topics to the most representative users. Ensuring that the co-staying network remains accessible to users of the RS is a non-trivial task. One mechanism could involve providing sufficient criteria to filter topics and users to avoid information overload. Furthermore, the proportion of the co-staying network that is visible to the user should also be controlled, for example by implementing a “zoom in / zoom out” design pattern.

Users should also be able to refocus the network based on their interests or goals. The system should allow seamless transition between vertices, for example between hotels, from a hotel to a topic, or from a topic to the users who referenced it in their reviews. Interacting with a topic should bring up the review snippets in which it is mentioned. The user could then expand reviews that look promising to read them fully; alternatively, less interesting snippets could be hidden completely.

Trust in online reviews has been shown to depend on the credibility of the source [19]. Thus, users should be afforded the possibility to explore the public profiles of reviewers that have contributed opinions about topics of interest. Public profiles might contain information about contributed reviews, visited hotels, frequently-mentioned topics, as well as any demographic data that users may want to share (e.g., purpose of travel, typical number of nights spent in hotels, average price paid). Trustworthiness cues embedded in the personal profiles could help users decide whether to consider—or not—the comments left by other users in the network regarding a certain recommendation. We expect that this might lead to a better calibration between the user’s trust in the RS and the system's actual trustworthiness [5].

DISCUSSION AND FUTURE WORK

Including hotel topics as an additional type of vertex in the network is expected to allow the RS to identify, with greater accuracy, which items users are likely to find attractive based on the attributes mentioned in their reviews as well as in reviews left by similar users [3,11]. At the same time, the system could extract and present, for each recommended item, the experiences of other people who are interested in the same combination of topics as the current user. The co-staying network might further facilitate the discovery of users with similar preferences and who have already stayed in the same hotel or in hotels with similar characteristics to the one that a person is considering. We believe that this could thus introduce novel ways of interacting with RS. For example, someone who has a strong preference for soft beds would be able to explore the opinions of other travelers who share this preference. Taking this a step further, the user might then ask herself, “What *other* preferences do such people have?” Personalizing the presentation to facilitate the exploration of similar people’s expectations could allow users to discover new preferences that they would not have considered otherwise. A prototype implementation of the presented co-staying concept and interactive mechanisms is currently under development.

We believe that our co-staying network concept has the potential to increase the transparency of the RS by providing visual clues as to why certain recommendations are presented. Moreover, the proposed interactive mechanisms are meant to increase user control over the presentation of the recommended items. Thus, the trustworthiness of the recommendation should increase [6]. Developing evaluation criteria for measuring the effects of these novel interactions on the trustworthiness of the recommendations is planned for future work.

Our approach should also alleviate, to some extent, the data sparsity problem by sharing topics—and, therefore, user opinions—about hotels that are part of a chain. However, this will continue to remain an issue in the case of isolated hotel vertices, for which not enough user-generated information is available from the network. This aspect will be investigated in future work.

ACKNOWLEDGMENTS

This work is supported by the German Research Foundation (DFG) under grant No. GRK 2167, Research Training Group "User-Centred Social Media".

REFERENCES

1. Barbu, C.-M. & Ziegler, J. User Model Dimensions for Personalizing the Presentation of Recommendations. In *Proc. IntRS '17*. ACM (to appear).
2. Borgatti, S. P. & Everett, M. G. Network analysis of 2-mode data. *Social networks 19*, 243–269 (1997).
3. Clements, M., de Vries, A. P., & Reinders, M. J. Optimizing single term queries using a personalized Markov random walk over the social graph. In *ECIR Workshop on Exploiting Semantic Annotations in Information Retrieval (ESAIR '08)*, Springer (2008).
4. Feuerbach, J., Loepp, B., Barbu, C.-M., & Ziegler, J. Enhancing an Interactive Recommendation System with Review-based Information Filtering. In *Proc. IntRS '17*. ACM (to appear).
5. Harman, J. L., O’Donovan, J., Abdelzaher, T., & Gonzalez, C. Dynamics of human trust in recommender systems. In *Proc. RecSys '14*, 305–308, ACM (2014).
6. Konstan, J. A. & Riedl, J. Recommender systems: From algorithms to user experience. *User Mod. and User-Adapted Interaction 22*, 1-2, 101–123 (2012).
7. Lambiotte, R. & and Ausloos, M. Collaborative tagging as a tripartite network. In *Proc. ICCS '06*, 1114–1117, Springer-Verlag Berlin Heidelberg (2006).
8. Linden, G., Smith, B., & York, J. Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet computing*, 7(1), 76–80. IEEE (2003).
9. Luong, N. T., Nguyen, T. T., Hwang, D., Lee, C. H., & Jung, J. J. Similarity-based Complex Publication Network Analytics for Recommending Potential Collaborations. *J.UCS*, 21(6), 871–889 (2015).
10. Maivizhi, R., Sendhilkumar, S., & Mahalakshmi, G. S. A Survey of Tools for Community Detection and

Mining in Social Networks. In *Proc. ICIA '16*, 71–79, ACM (2016).

11. McAuley, J. & Leskovec, J. Hidden factors and hidden topics: Understanding rating dimensions with review text. In *Proc. RecSys '13*, 165–172, ACM (2013).
12. Rossetti, M., Stella, F., & Zanker, M. Analyzing user reviews in tourism with topic models. *Information Technology & Tourism*, 16(1), 5–21 (2016).
13. Sacha, D., Senaratne, H., Kwon, B. C., Ellis, G., & Keim, D. A. The role of uncertainty, awareness, and trust in visual analytics. *IEEE Trans. Vis. Comp. Graph.*, 22-1, 240–249 (2016).
14. Shavitt, Y. & Weinsberg, U. Song clustering using peer-to-peer co-occurrences. In *Proc. ISM '09*. 11th IEEE International Symposium on Multimedia, 471–476, IEEE (2009).
15. Sen, S., Vig, J., & Riedl, J. Tagommenders: connecting users to items through tags. In *Proc. WWW '*09, 671–680, ACM (2009).
16. Verbert, K., Parra, D., Brusilovsky, P., & Duval, E. Visualizing recommendations to support exploration, transparency and controllability. In *Proc. IUI '13*, 351–362, ACM (2013).
17. Wasserman, S. & Faust, K. Social Network Analysis: Methods and Applications. Cambridge Univ. Press (1994).
18. Xu, G., Wu, Z., Zhang, Y., & Cao, J. Social networking meets recommender systems: survey. *International Journal of Social Network Mining*, 2(1), 64–100 (2015).
19. Yoo, K. H., Lee, Y., Gretzel, U., & Fesenmaier, D. R. Trust in travel-related consumer generated media. *Information and communication technologies in tourism 2009*, 49–59, Springer (2009).
20. Zhou, D., Zhu, S., Yu, K., Song, X., Tseng, B. L., Zha, H., & Giles, C. L. Learning multiple graphs for document recommendations. In *Proc. WWW '08*, 141–150. ACM (2008).

