# Toward a Design Space for the Recommendation of Communities

**Sven Buschbeck**
German Research Center for Artificial
Intelligence

**Anthony Jameson**
German Research Center for Artificial
Intelligence

## ABSTRACT

Recent years have seen an explosion in the number of on-line communities; a person looking for a community for some purpose may have hundreds to choose from. Helping users make such choices therefore constitutes an interesting application area for recommender systems, which has so far been explored only sporadically. To facilitate further exploration of this area, we discuss the following questions: What goals can a user be pursuing when looking for a community? What sorts of information about communities can be available that could be used by a recommender system? The answers to the second question are based on a systematic examination of a sample of 15 communities on three web-based platforms.

## 1 INTRODUCTION

One reason for the vitality of the recommender systems field is the great variety of types of entities that can be recommended. Each new type of entity may have a different set of relevant properties, and people may have different reasons for being interested in that type of entity. Therefore, new combinations of recommendation techniques and interfaces may be required.

A case in point is the class of online communities, whose importance has been growing rapidly in recent years. Many people now spend a considerable proportion of their time acting as a member of one online community or the other; and they may have more superficial contact with a larger number of other online communities. Consequently, helping people to find the right online community for a given purpose is a promising application area for the recommender systems community.

This short paper aims to strengthen the existing foundation for studying the recommendation of communities by briefly summarizing previous work on this topic; considering the various goals that people may have when looking for an online community; and surveying the various properties of online communities that may be relevant to recommendation. The discussion of this latter point is fleshed out with an analysis of a sample of web-based communities.

Because the authors are especially interested in online communities for sharing media, parts of the discussion are especially relevant to these and similar communities.

## 2 RELATED WORK

We use the term *community* to denote a collection of computer users who have joined an explicitly defined group on a platform (such as Orkut, Flickr, or Facebook) that supports the formation of such groups and provides facilities for the members to exchange information with each other and engage in other joint activities. In some platforms, other terms such as *group* are used instead. Hence, in our terminology a web portal like Flickr is referred to as platform or a network of communities rather than as a community.

Brocco and Groh ([1]) discuss the problem of *team recommendation*, introducing a useful set of general concepts for characterizing both team members and teams. Since their focus is on composing new teams rather than recommending existing ones, and since working teams are different in some ways from online communities in general, the results of this work require some adaptation to be applied to the problem of community recommendation.

The problem of helping users to find interesting communities was addressed directly by the TOUCHGRAPH LIVEJOURNAL BROWSER (see, e.g., [6]), which uses information visualization techniques to show a user what communities of interest his or her friends belong to. Although this sort of visualization does not constitute a recommendation, it could be used as part of a recommender system for communities.

Similarly, Diedrich et al. focused in [5] on creating user profiles on the basis of the tagging behavior of each user. Although this approach does not deal directly with the recommendation of communities to users, it does provide a promising way of classifying users to whom a given community might be recommended.

Information about the communities that friends belong to, along with other information, was used for real recommendation of communities in the system SONARS ([2]).

Spertus et al. ([7]) empirically explored methods for recommending communities to a given user on the basis of their similarity to a community that the user belongs to. They compared several similarity metrics that consider only the set of members of each community, treating each commu-

nity as a *bag of users*. Chen et al. ([4]) went a step further by treating each community in their algorithms as both a bag of users and a *bag of words* (e.g., a semantic description of the community). As we will discuss below, there is a great deal of other information about communities that a recommender could in principle take into account, but focusing on bags of users and/or bags of words has advantages in terms of efficiency and scalability, and it remains to be seen what can be gained by taking other types of information into account.

Research by Chen et al. [3] yields some relevant insights even though it concerns the recommendation of individual persons rather than the recommendation of communities. Comparing several person recommendation algorithms, these researchers found that different algorithms and types of information were best suited to different goals and contexts (e.g., looking for people that the user already knows vs. looking for new friends).

Our brief analysis in this paper argues that this idea is likely to apply to community recommendation as well: We will point to the diverse goals that people can have when looking for communities and to the diverse types of information about communities that are commonly available. Our aim is to motivate research on the relevance of different types of information for recommendation to users with different goals.

## 3 POSSIBLE USER GOALS

There are many reasons why a user might be interested in finding a community, some less obvious than others. Even though a recommender system may not be able to find out much about the current user's goals, it is important for designers to be aware of this range of goals. Since an exhaustive analysis of users' possible goals would be beyond the scope of this paper, we present several examples to illustrate the range of possibilities.

Note in particular that a user may be able to benefit from the existence of a given community without joining that community or even interacting with its members. Regardless of their primary reason for existence, communities can be exploited as information resources.

**Goal: Find Information or Media:** A user may want to get information or media relating to a specific topic. For example, a journalist may be searching for photos of erupting volcanoes while writing a story about the recent eruption in Iceland. If there is an active community of experts on that topic, it might be a good option to join that community or at least try to get access to the collection of media maintained by that community. Note that in some cases it may be easier to find a community of people interested in a given topic than to find a specific piece of information or a particular medium related to that topic. So even if the community is not in itself of interest to the user, it can serve as a stepping stone to the desired information or media.

**Goal: Establish Real-World Contacts:** A user may want to organize some kind of event, such as a local soccer match for children, and need to acquire helpers. Hence she is searching for people with specifically required knowledge, skills and experience from her region willing to support her. A community of people from the same geographical region who are interested in similar topics may be a good place for her to find the required persons.

**Goal: Get On-Line Help and Feedback:** A user may be working on a specific task and be looking for help and/or feedback. As with the previous goal, a relevant online community can be a useful resource; but the requirements placed on such a community are partly different.

**Goal: Offer Information or Media:** Another goal could be to offer information or media to other users. For example, suppose that a user is a member of an open source software project. Now that a new major release is coming up, she would like to inform other potentially interested users about the new features and invite them to download and test the new piece of software. Especially helpful to the user will be large communities interested in this topic. Additionally, the user may prefer communities with highly active members who are likely to respond promptly.

## 4 ANALYSIS OF TYPICALLY AVAILABLE INFORMATION

Finding a community that satisfies a particular goal requires taking into account relevant information about that community—or having a recommender system take it into account. Since communities are complex entities, they have many qualitatively different attributes (as has been discussed in some of the related work summarized in section 2). But which of these attributes might be known to a recommender system, and what goals might they be relevant to?

As an initial empirical basis for answers to these questions, we analyzed a sample of communities on three widely used web platforms: Orkut (which supports social networks in general, without restriction to a particular population or type of content), Flickr (which supports communities that exchange visual media), and LinkedIn (which supports business-related communities). These three platforms were chosen because each is typical of a particular type of social network: generic, media-related, and business-related social networks, respectively. For each platform, 5 communities of 1000 to 40000 users (14000 users on the average) from 5 popular topic areas (technology, soccer, news, travel, and nature) were selected for examination.

Table 2 discusses the information that could be found about a community's individual members, while Table 1 looks at the attributes of the communities themselves. The tables also include remarks about how the various types of information about communities might be useful for the recommendation of communities to users with particular goals.

| Question Considered | | Detailed Results | | | Overall Results | |
|---|---|---|---|---|---|---|
| *Question Category* | *Question* | Orkut | Flickr | LinkedIn | To what extent was this information found in our sample of communities on three platforms? | How could this information be useful in the recommendation of communities to users with particular goals? |
| Demographic Variables | What is their age distribution? | Unknown | Unknown | Unknown | Unknown | Users searching for social contact might be interested in a specific age range. |
| | Do we know the gender distribution? | Yes | Maybe (Optional user profile field) | No | Mandatory field in user profiles on Orkut, optional on Flickr, non-existing on LinkedIn. | If a user is interested in finding contacts to meet in person or to talk about gender-specific topics, he or she might want to filter by gender |
| | What languages do the users speak? | Mostly English | Mostly English | Mostly English | Mostly available | A user's languages must in general include the community's language in order that the user can make any use of the community (except for some superficial uses) |
| | Can we get to know the geographical location of the user? | Yes | Maybe | Maybe (Location might be derived from user's occupation) | Mostly available; country has to be entered in Orkut and users of Linkedin tell their location indirectly by naming their current employer, but it is an optional property on Flickr. | If a user's intention is to meet community members personally, the geographical location is of very high importance. |
| Fuctional variables | Can we get to know something about skills, interests or hobbies? | Maybe (optional list of interests) | Maybe (Only as part of user´s plain text description) | Yes (skills and education) | Always available on business-oriented platforms, less often on social networks | If a user wants to hire / acquire people, skills and eduction are very important to make a decission |
| Social network variables | Can we get to know something about other connections of this user? | Yes | Yes | Yes | | |
| | Can we access the number of a user's contacts and communities, and further about these? | Yes | Yes | Yes | Always available | If the searcher aims to establish contact, another user's contacts and communities can be an indicator of the how much the other person is interested in establishing new contacts as well |

**Table 1.** *Analysis of recommendation-relevant information about the members of online communities that can be found in a sample of communities on three platforms.*

# 5 QUESTIONS FOR FUTURE RESEARCH

The ideas and information presented above and in the tables are intended to serve as a starting point for the investigation of issues like the following:

- What recommendation techniques are best suited for matching users who have particular goals with communities of a given type?
  Although it seems unlikely that every combination of user goals and community type will call for a different recommendation algorithm, it also seems unlikely that a single community recommender system can perform effectively in all of the situations mentioned above.
- What particular forms can interaction with a community recommender system take?
  For example, to what extent should the system try to acquire information about the user's goals in looking for a community, and how could it do so?

Though we have taken but a first small step toward answering these questions, we believe that it is important to start with a realistic understanding of the range of user goals and the types of available information.

# REFERENCES

1. Michele Brocco and George Groh. Team recommendation in open innovation networks. In *Proceedings of the 2009 ACM Conference on Recommender Systems*, pages 365–368, New York, 2009. ACM.

2. Francesca Carmagnola, Fabiana Vernero, and Pierluigi Grillo. SoNARS: A social networks-based algorithm for social recommender systems. In *Proceedings of UMAP 2009, the 17th International Conference on User Modeling, Adaptation, and Personalization*, Berlin, 2009. Springer.

3. Jilin Chen, Werner Geyer, Casey Dugan, Michael Muller, and Ido Guy. "Make new friends, but keep the old" – Recommending people on social networking sites. In Saul Greenberg, Scott Hudson, Ken Hinckley, Meredith R. Morris, and Dan R. Olsen, editors, *Human Factors in Computing Systems: CHI 2009 Conference Proceedings*, pages 201–210. ACM, New York, 2009.

4. Wen-Yen Chen, Dong Zhang, and Edward Y. Chang. Combinational collaborative filtering for personalized community recommendation. In *Proceedings of KDD 2008*, New York, 2008. ACM.

5. Jrg Diederich and Tereza Iofciu. Finding communities of practice from user profiles based on folksonomies. In *First International Workshop on Building Technology-Enhanced Learning Solutions for Communities of Practice*, Crete, 2006.

6. Amy Soller. Adaptive support for distributed collaboration. In Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl, editors, *The Adaptive Web: Methods and Strategies of Web Personalization*, pages 573–595. Springer, Berlin, 2007.

7. Ellen Spertus, Mehran Sahami, and Orkut Buyukkokten. Evaluating similarity measures: A large-scale study in the Orkut social network. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, pages 678–684, New York, 2005. ACM.

| Question Considered | | Platform-Specific Results | | | Overall Results | |
|---|---|---|---|---|---|---|
| *Question Category* | *Question* | Orkut | Flickr | LinkedIn | To what extent was this information found in our sample of communities on three platforms? | How could this information be useful in the recommendation of communities to users with particular goals? |
| Group Longevity | Do we know how long the community has existed? | Yes | Yes | Yes | Always | In combination with the community's activity, it can be an indicator of the quality of organization and motivation of community members. |
| Organization | Is there someone who is in charge? | Yes | Yes | Yes | The creator becomes the owner of a community | The owner's user profile can be an indicator of the community's intended purpose / topic. |
| | Can we learn something about the geographical location of the community? | Yes: mostly Asia and South America | In some cases | In some cases | There is a geographical location for each community in Orkut only; in the other networks, the community location could be calculated from the location of the members. | If a user is interested in meeting the members in person, geographical proximity is of great interest. It also has implications concerning other likely properties of members. |
| Social Performance | What is the number of community members? | 13666 on the average | 16154 on the average | 13316 on the average | Always available | The number of users represents the popularity of the community's topic, but it can also be seen as indicator of the quality of the content and organization. If a user is looking for feedback, the number of community members is important. |
| Information System | Does the community maintain some collection of objects? | Yes | Yes | Yes | All communities maintain a collection of some kind of objects | The collected items (regardless of their type) are a good source for knowing how active a community is and what it is interested in. |
| | If Yes, what sort of objects are maintained in the collection? | Events and surveys | Media | News items, job offers | On two of three platforms, communities maintain a collection of events (also called "News") | |
| | If it is a collection of media items, what kind of media? | n/a | Photos and videos | n/a | Supported types of media are always available in communities with a media collection | Image features can be extracted to learn about the interests of the community. |
| | What is known about the media? | n/a | Title, description, tags, and device meta data (EXIF) | n/a | A lot of detailed information | Metadata usually contains location, date, and topic in terms of title, description, and tags, thus providing insight into the community's interests. |
| | How are the media items geographically distributed? | n/a | Mostly Europe, North America, and Central Asia | n/a | Locations can be plotted on a map if geographic data is available. | Relevant if a user is interested in a specific region or if she is interested in a specific topic and wants to be sure that the content is not dominated by a particular local point of view. |
| | How many items are in the collection? | 23 surveys and 0 events on the average | 29,3311 on the average | Mostly unknown | Mostly available | On the one hand, a high number of items can be evidence of the community's motivation, but it can also mean that the community is less organized or focused. |
| | What is the source of the items? | Members of C | Members of C | Members of C | Always available | The number of contributions from a given user can be used as a weight for the corresponding user profile when the system tries to learn about the community's interests. |
| | Can members comment on objects? | Yes | Yes | Yes | Always available | The number of comments suggests how active a community is and how much the members are interested in social interaction. |
| Communication System | Can we learn something about the message-based communication between the members of C? | Yes | Yes | Yes | Mostly available; sometimes only to community members. | (See the comments below) |
| | If Yes, is there a forum? | Yes | Yes | Yes | Always available | |
| | If Yes, how many posts does the forum contain? | 229 on the average | 942 on the average | Mostly unknown | | The number of posts relative to the number of community members can be evidence of the quality of content and moderation, and focus of members' interests. |
| | How many replies are there to each post? | 7 on the average | 34 on the average | Mostly unknown | Indirectly available by counting replies. High variability: Some posts got 100+ replies, while many remained unanswered. | The ratio of number of replies to number of posts can represent the users' willingness to communicate and interact. |
| | Is there a specific language associated with the community? | Yes | No | No | The majority of communities do not define the spoken language in advance, but English is the language of the vast majority of communities. | User's language must match community in order that the user can make any use of the community. |
| | Are there communities with mixed languages? | No | No | No | None of these communities contained discussions in different languages. | If a community's language is not given explicitly, it can be derived from the members' communication |
| | Can users communicate with each other directly? | Yes | Yes | Yes | Means of direct communication are available, but the communications themselves are not accessible from the outside for privacy reasons | Users' current interests could be derived from communication content if it was available. |
| | If Yes, how? | Mail, instant messaging | Mail | Mail | At least email is implemented on all platforms. | Communication frequency can be an indicator of how communicative a user is. |

**Table 2.** *Analysis of recommendation-relevant information about online communities as such that can be found in a sample of communities on three platforms.*

4