Chapter 1

Investigating the Properties of a Social Bookmarking and Tagging Network

Ralitsa Angelova
Max Planck Institut für Informatik, Germany

Marek Lipczak
Dalhousie University, Canada

Evangelos Milios
Dalhousie University, Canada

Pawel Pralat
Dalhousie University, Canada

ABSTRACT

Social networks and collaborative tagging systems are rapidly gaining popularity as a primary means for storing and sharing data among friends, family, colleagues, or perfect strangers as long as they have common interests. del.icio.us is a social network where people store and share their personal bookmarks. Most importantly, users tag their bookmarks for ease of information dissemination and later look up. However, it is the friendship links, that make del.icio.us a social network. They exist independently of the set of bookmarks that belong to the users and have no relation to the tags typically assigned to the bookmarks. To study the interaction among users, the strength of the existing links and their hidden meaning, we introduce implicit links in the network. These links connect only highly “similar” users. Here, similarity can reflect different aspects of the user’s profile that makes her similar to any other user; such as number of shared bookmarks, or similarity of their tags clouds. The authors investigate the question whether friends have common interests, they gain additional insights on the strategies that users use to assign tags to their bookmarks, and they demonstrate that the graphs formed by implicit links have unique properties differing from binomial random graphs or random graphs with an expected power-law degree distribution.

DOI: 10.4018/978-1-61350-474-1.ch001
INTRODUCTION

Social bookmarking and collaborative tagging services lead to the formation of a new type of organically grown network structure. In such networks, users are linked to other users through social connections (e.g. directed friendship links) and to network specific online resources (e.g. bookmarks, photos, books, etc.) by either explicitly linking to them, tagging them with appropriate terms, or commenting on them. Clustering users in this context is a challenging problem, as it involves accounting for multiple types of social linkage among users and diversity of content ranging from personal photos (flickr.com) or bookmarks (del.icio.us) to whole libraries of read books throughout the user’s lifetime (The Personal Library, librarything.com). The complexity of the clustering problem raises dramatically if we look at the overall electronic fingerprint of these users after connecting all their profiles from the various social networks they actively contribute to (Moser et al., 2007). Not only is clustering itself challenging but evaluation of the clustering solution is also very hard as reference class assignments are typically missing or very expensive to manually gather. These class assignments (also known as ground truth) are ignored in the clustering process. They are used exclusively in the evaluation phase to compare the groups produced by the clustering technique to the known classes it comprises. Modelling social bookmarking and tagging services is a way to generate synthetic data sets that mimic the behaviour of such social networks. Moreover, the synthetic data generative model also provides the corresponding ground truth for performance evaluation and comparison. A requirement for the design of useful models is an in-depth understanding of the properties of real-life data sets obtained from on-line social networks such as del.icio.us.

A collaborative tagging system like del.icio.us can be visualized as a tripartite structure (Halpin et al., 2007), where links (edges) are established between users, tags and bookmarks. Additionally, the social dimension introduces “friendship” links between users. Several research questions about the structure of the social network and its implications arise:

- What is the role of friendship in relation to interest sharing as reflected in the bookmarks and tags of users. Do friends appear to have more common interests than non-friends?
- Do “highly social” users share more topics of interest with others than the “less social” users?
- What are the structural properties of the friendship graph and the graphs induced by the implicit similarity-based links among users? Is their degree distribution indicative of power-law graphs? What are their connectivity and local density properties, measured by the clustering coefficient, as a function of in their -core analysis (Healy et al., 2008)? How do they compare with binomial random graphs (Janson et al., 2000) and random graphs with an expected power-law degree distribution (Chung and Lu, 2004)?
- What are the common properties of the friendship, bookmark-based and tag-based links? In particular, how do the three types of links correlate for individual users?

We are not the first to analyze social collaboration on the Web. Evolution models of two online social networks - Flickr and Yahoo! 360 are examined in (Kumar et al., 2006). In the experiments performed in (Negoeescu, 2007) on the photo sharing network Flickr, after taking a random subset of photos and their owners or users, it is demonstrated that Flickr exhibits the characteristics of small-world and scale-free networks described earlier by (Barabasi, 2002, Cohen et al., 2003). Search and ranking techniques applied to social networks are discussed in (Hotho et al., 2006).