Social Search in Collaborative Tagging Networks: The Role of Ties

Von der Fakultät für Elektrotechnik und Informatik
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des Grades
DOKTORIN DER NATURWISSENSCHAFTEN
Dr. rer. nat.
genehmigte Dissertation

von

Frau Kerstin Bischoff, M. A.
geboren am 1980/04/02, in Lauchhammer

2013

Referent: Prof. Dr. techn. Wolfgang Nejdl
Korreferent: Prof. Dr. Michael Rohs

Tag der Promotion: 2013/06/25

Information is not knowledge
Knowledge is not wisdom
Wisdom is not truth
Truth is not beauty
Beauty is not love
Love is not music
Music is the best

Frank Zappa - Packard Goose

Zusammenfassung

Soziale Netzwerke im Internet sind heutzutage sehr populär. Auf Plattformen wie *Facebook* oder *Google+* vernetzen sich Menschen online mit ihren guten Freunden und alten Bekannten oder sie befreunden sich neu mit bisher nicht gekannten Personen. Ebenso entstehen neue (indirekte) Beziehungen durch das Lesen und Übernehmen aus anderer Leute Blogs oder Tweets. Und in Systemen zum gemeinschaftlichen Verschlagworten (Collaborative Tagging Systems), wie beispielsweise *Last.fm*, *Flickr* oder *Delicious*, teilen Internetnutzer ihre Lesezeichen und mit Tags verschlagwortete Web-Ressourcen mit Freunden sowie unbekannten, ähnlich interessierten Nutzern. Dabei bieten diese verschiedenen Arten von Beziehungen unterschiedliche Potenziale im Hinblick auf Informationsaustausch und Zusammenarbeit. Die vorliegende Arbeit widmet sich sozialen Beziehungen in Tagging-Systemen mit dem Ziel, erste Voraussetzungen für ihre erfolgreiche Verwertung in Techniken zur 'sozialen' Suche in beziehungsweise aufbauend auf solchen Systemen zu schaffen.

Zuerst betrachten wir soziale Annotationen (Tags) und ihren Mehrwert für Suchstrategien auch ohne explizit gegebene Freundschaftsbeziehungen, zum Beispiel durch adäquatere Benutzermodellierung, umfassendere Beschreibung von Ressourcen, via Text Mining neu gewonnenes Wissen oder auch das Finden von gleichgesinnten Personen oder Experten. Wir analysieren daher ausführlich die Arten von Tags und ihre Häufigkeiten in verschiedenen Systemen sowie die sich daraus ergebenden Implikationen für Such- und Empfehlungssysteme. Aufbauend auf den empirischen Ergebnissen präsentiert diese Arbeit dann Ansätze zur automatischen Anreicherung von Ressourcen und Benutzerprofilen mit zusätzlichen Informationen – Themen und Stimmungen von Musikstücken. In einem zweiten Teil erfolgt die Analyse vorhandener Freundschaftsbeziehungen im Musikportal *Last.fm*. Dabei werden Online- und "reale" Offline-Beziehungen gegenüberstellend verglichen. Hier untersuchen wir vor allem Ähnlichkeiten zwischen Freunden im Hinblick auf demografische Daten, lokale Netzwerkstruktur und ganz besonders Musikgeschmack. Während in *Google+* und *Facebook* Benutzer Freundeskreise oder -listen noch manuell verwalten müssen, zeigen wir, wie maschinelle Lernverfahren genutzt werden können, um Online- und Offline-Freunde automatisch zu identifizieren. Weitere Experimente mit *Wikipedia*-Daten bestätigen, dass auch die Vorhersage zukünftigen Verhaltens von der Berücksichtigung sozialer Beziehungen profitiert.

Schlagwörter: *Soziale Netzwerke, Tagging, Beziehungsstärke, Maschinelles Lernen*

Abstract

With the rise of the Web 2.0 online social networking has become a huge trend. On *Facebook*, *Google+*, etc. people connect with their friends or make new friends. They form new (indirect) connections by reading and adopting from other peoples' blogs or tweets. Similarly, in collaborative tagging systems like *Last.fm*, *Delicious*, or *Flickr* people share bookmarks and tagged resources with friends or unknown, like-minded users. Just like in many off-line situations, it has been argued that the different kinds of online ties hold different potentials for information exchange and collaboration. In this work we study social ties in collaborative tagging systems – a prerequisite for successfully exploiting the different kinds of ties within social search in and built upon such systems.

First, we focus on social tags as means for enhancing search even without explicitly given social connections: through better user profiling, richer resource descriptions, newly mined knowledge, or the recommendation of people with similar interests. In order to prove that tags are indeed a useful source of additional information, we analyze tag usage patterns in diverse tagging systems and discuss the implications for user profiling, search, and recommendation. Building upon the found characteristics we present approaches exploiting tags to enrich resources or user profiles with additional information – music moods and themes. Second, we examine existing friendship links in *Last.fm* contrasting online and "real-world" friends having co-attended events. We investigate in depth similarity along such social links regarding demographics, network structure as well as taste in particular. While in platforms like *Google+* or *Facebook* users still have to manually maintain circles or lists of close friends, family, etc., we are developing machine learning methods that successfully identify online and off-line friends of different strength automatically. Additional experiments on weak and strong ties in *Wikipedia* show that also the prediction of future behavior, here co-editing of articles, can benefit from considering social ties.

Keywords: *Social networks, collaborative tagging, tie strength, machine learning*

Contents

List of Figures

List of Tables

Preface

Most of the work covered in this thesis has been published in individual papers.

Preliminary ideas for this work appeared in:

- Bischoff, K. Exploiting Social Ties for Search and Recommendation in Online Social Networks - Challenges and Chances. In *Proceedings des 22. Workshop über Grundlagen von Datenbanken des GI-Arbeitskreises Grundlagen von Informationssystemen* (Bad Helmstedt, Germany, May 25-28 2010), GvDB '10, CEUR Workshop Proceedings (CEUR-WS.org)

The usage study reported in Chapter 2 is based on:

- Anjorin, M., Rensing, C., Bischoff, K., Bogner, C., Lehmann, L., Reger, A. L., Faltin, N., Steinacker, A., Lüdemann, A., and Domínguez García, R. CROKODIL - A Platform for Collaborative Resource-Based Learning. In *Proceedings of the 6th European Conference of Technology Enhanced Learning* (Palermo, Italy, September 20-23 2011), EC-TEL '11, Springer-Verlag, pp. 29–42

Chapter 4 builds upon:

- Bischoff, K., Firan, C. S., Nejdl, W., and Paiu, R. Can All Tags be Used for Search? In *Proceedings of the 17th International ACM Conference on Information and Knowledge Management* (Napa Valley, California, USA, October 26-30 2008), CIKM '08, ACM, pp. 193–202
- Bischoff, K., Firan, C. S., Paiu, R., Nejdl, W., Laurier, C., and Sordo, M. Music Mood and Theme Classification - A Hybrid Approach. In *Proceedings of the 10th International Society for Music Information Retrieval Conference* (Kobe, Japan, October 26-30 2009), ISMIR '09, International Society for Music Information Retrieval, pp. 657–662

Chapter 4 also references work published in:

- Bischoff, K., Firan, C. S., Kadar, C., Nejdl, W., and Paiu, R. Automatically Identifying Tag Types. In *Proceedings of the 5th International Conference on Advanced Data Mining and Applications* (Beijing, China, August 17-19 2009), ADMA '09, Springer-Verlag, pp. 31–42
- Bischoff, K., Firan, C. S., Georgescu, M., Nejdl, W., and Paiu, R. Social Knowledge-Driven Music Hit Prediction. In *Proceedings of the 5th International Conference on Advanced Data Mining and Applications* (Beijing, China, August 17-19 2009), ADMA '09, Springer-Verlag, pp. 43–54
- Iofciu, T., Fankhauser, P., Abel, F., and Bischoff, K. Identifying Users Across Social Tagging Systems. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media* (Barcelona, Catalonia, Spain, July 17-21 2011), ICWSM '11, AAAI Press, pp. 522–525
- Bischoff, K., Firan, C. S., Nejdl, W., and Paiu, R. Bridging the Gap Between Tagging and Querying Vocabularies: Analyses and Applications for Enhancing Multimedia IR. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web 8*, 2-3 (July 2010), 97–109
- Bischoff, K., Firan, C. S., Nejdl, W., and Paiu, R. How Do You Feel About "Dancing Queen"?: Deriving Mood & Theme Annotations from User Tags. In *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital libraries* (Austin, Texas, USA, June 15-19 2009), JCDL '09, ACM, pp. 285–294
- Bischoff, K., Firan, C. S., and Paiu, R. Deriving Music Theme Annotations from User Tags. In *Proceedings of the 18th International Conference on World Wide Web* (Madrid, Spain, April 20-24 2009), WWW '09, ACM, pp. 1193–1194

The work on identifying weak and strong ties in *Last.fm* reported in Chapter 5 was published in:

- Bischoff, K. We Love Rock 'n' Roll: Analyzing and Predicting Friendship Links in Last.fm. In *Proceedings of the ACM Web Science Conference* (Evanston, Illinois, USA, June 22-24 2012), WebSci '12, ACM, pp. 47–56. `http://doi.acm.org/10.1145/2380718.2380725` (© ACM, 2012)

Other publications not mentioned in this thesis due to space limitations are:

- Tran, N. K., Zerr, S., Bischoff, K., Niederee, C., and Krestel, R. Topic cropping: Leveraging latent topics for the analysis of small corpora. In *Proceedings of the 17th International Conference on Theory and Practice of Digital Libraries* (Valletta, Malta, September 22-26 2013), TPDL 2013, Springer-Verlag
- Andrews, P., De Natale, F., Buschbeck, S., Jameson, A., Bischoff, K., Firan, C. S., Niederée, C., Mezaris, V., Nikolopoulos, S., Murdock,

V., and Rae, A. GLOCAL: Event-based Retrieval of Networked Media. In *Proceedings of the 21st International Conference Companion on World Wide Web* (Lyon, France, April 16-20 2012), WWW '12 Companion, ACM, pp. 219–222

- Zerr, S., Bischoff, K., and Chernov, S. GuideMe! The World of Sights in Your Pocket. In *Proceedings of the 27th International IEEE Conference on Data Engineering* (Hannover, Germany, April 11-16 2011), ICDE '11, IEEE Computer Society, pp. 1348–1351
- Bischoff, K., Herder, E., and Nejdl, W. Workplace Learning: How We Keep Track of Relevant Information. In *Proceedings of the 2nd European Conference on Technology Enhanced Learning* (Crete, Greece, September 17-20 2007), EC-TEL '07, Springer-Verlag, pp. 438–443
- Bischoff, K., Mandl, T., Kölle, R., and Womser-Hacker, C. Geographische Bedingungen im Information Retrieval: Neue Ansätze in Systementwicklung und Evaluierung. In *Proceedings of the 10th International Symposium for Information Science* (Cologne, Germany, May 30-June 1 2007), ISI '07, UVK Verlagsgesellschaft mbH, Konstanz, Germany, pp. 15–26
- Gey, F., Larson, R., Sanderson, M., Bischoff, K., Mandl, T., Womser-Hacker, C., Santos, D., Rocha, P., Di Nunzio, G. M., and Ferro, N. Challenges to Evaluation of Multilingual Geographic Information Retrieval in GeoCLEF. In *Proceedings of the 1st International Workshop on Evaluating Information Access* (Tokyo, Japan, May 15 2007), EVIA '07, `http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings6/EVIA`, pp. 74–77
- Gey, F. C., Larson, R. R., Sanderson, M., Bischoff, K., Mandl, T., Womser-Hacker, C., Santos, D., Rocha, P., Di Nunzio, G. M., and Ferro, N. GeoCLEF 2006: The CLEF 2006 Cross-Language Geographic Information Retrieval Track Overview. In *Proceedings of the 7th Workshop of the Cross-Language Evaluation Forum* (Alicante, Spain, September 20-22 2006), CLEF '06, Springer-Verlag, pp. 852–876
- Bischoff, K., Mandl, T., and Womser-Hacker, C. Blind Relevance Feedback and Named Entity Based Query Expansion for Geographic Retrieval at GeoCLEF 2006. In *Proceedings of the 7th Workshop of the Cross-Language Evaluation Forum* (Alicante, Spain, September 20-22 2006), CLEF '06, Springer-Verlag, pp. 946–953
- Bischoff, K., Mandl, T., and Womser-Hacker, C. GeoCLEF 2006: Crosslinguales geographisches Information Retrieval. In *Proceedings of the German Society for Informatics (GI) Joint Workshop Event Lernen - Wissensentdeckung - Adaptivität, LWA 2006* (Hildesheim, Germany, October 9-11 2006), vol. 1/2006 of *Hildesheimer Informatik-Berichte*, University of Hildesheim, Institute of Computer Science, pp. 89–93

1 Introduction

With the advent of the Web 2.0 online social networking has become a huge trend. On platforms like *Facebook*[1] or *Google+*[2] people connect with their friends or make new friends. They form new (indirect) connections by reading and adopting from other peoples' blogs or tweets. Similarly, in the more purpose-oriented collaborative tagging systems like *Last.fm*[3], *Delicious*[4], or *Flickr*[5] people share bookmarks and tagged resources with friends or unknown, like-minded users.

Consequently, social search has received a lot more attention recently in research as well as by popular search engines like *Google*[6] or *Bing*[7]. For example, in 2010 Google bought Aardvark[8], a social search engine for question answering exploiting the user's social graph combined with relevance or expertise matching [110]. Social search can be defined as "search acts that make use of social interactions with others. These interactions may be explicit or implicit, co-located or remote, synchronous or asynchronous"[71]. In the model of social search proposed in [71] information exchange between people happens at various stages in the search process: before search to gather requirements and to formulate the representation of an information need, during search while information foraging and sensemaking, and, finally, after search when distributing search products to (close) others [71].

For instance, Collaborative Filtering systems use activities and opinions of (unknown) users for recommending information or products and as such can be considered "social search"[71]. *Amazon*[9] is a prominent example for such techniques. Similarly, social tags enrich web resources with human generated labels other users

[1] `http://www.facebook.com`
[2] `http://plus.google.com`
[3] `http://last.fm`
[4] `http://delicious.com`
[5] `http://www.flickr.com`
[6] `http://www.google.com`
[7] `http://www.bing.com`
[8] `http://techcrunch.com/2010/02/11/google-acquires-aardvark-for-50-million`
[9] `http://www.amazon.com`

of the system can build upon when browsing or searching information. Tags are also indicators of shared interest, indirectly connecting users, and thus potentially suitable for user profiling and matching. Of course, big part of social search is concerned with making use of one's social network. For example, a study on questions posted in *Facebook* status messages [165] showed that people turn to their online contacts especially for subjective queries asking for recommendations (29%) and opinions (22%), e.g., on restaurants (see also [110, 166]). 17% of the questions were factual in nature. Trust and personalization were found the main motivations for asking one's contacts; altruism and expertise for answering.

However, people in that study did not feel well with posting and answering highly private questions (e.g., dating or religion). One reason may be the big audience in online networking services like *Facebook*. Due to the ease of friending, links in online social networks are often spurious, shuffling together close friends and loose acquaintances all as “friends”. danah boyd [60] called this the “collapsing of context”.

In a study on the usefulness of social annotations on search results [166], users reported that such references to a personal contact who interacted with the web page are useful if they come from close persons or persons with known expertise on a topic. Other studies indicate as well that especially in taste domains, like music or books, familiarity, i.e. a strong tie, with the person generating the recommendation is appreciated by users and may lead to more precise recommendations [34, 200]. The importance of not treating all online relationships as equal has just recently been accounted for by introducing social circles in *Google+* and friend lists in *Facebook*.

In ‘real’ social networks, strong ties (i.e., family, close friends) and weak ties (loose acquaintances) have been found to show different characteristics. Weak ties are often ‘bridges’ connecting different communities, thus bringing new information (e.g. job seeking). Strong ties offer mutual support and trust, but they likely share knowledge, preferences, values, and friends [92, 93]. Regarding social search, this implies that the different kinds of ties hold different potentials with respect to diversity and novelty on the one hand and completeness and trust on the other hand. As McAfee [158] pointed out, different kinds of ties – if supported by the right technology – offer different potential benefits for information exchange and collaboration (from [158]):

- none/absent ties: collective intelligence (prediction market)
- potential ties: efficient search, tie formation (blogosphere)
- weak ties: innovation, non-redundant information, network bridging (social networking software)
- strong ties: collaboration, productivity, agility (wiki)

Though McAfee gives only one example technology for each kind of tie, the list can be easily extended. For example, CVS systems for cooperative file management with version control or cooperative workspaces with features like group chat, etc., are other tools supporting collaboration in small groups of strong ties. With our focus on collaborative tagging systems, we argue that the statistical patterns found in tagging systems exhibit the so called "wisdom of the crowds", beneficial for mining new knowledge and enhancing search even without any social connections between users. On the other hand, social tags can be used for user profiling, drawing an implicit or potential link between users of the same tag, thereby indicating similar interests. In addition, as most tagging systems support the social feature of friending, we find actual strong and weak ties usually not differentiated explicitly in the system.

As a prerequisite for successfully exploiting the different kinds of ties within social search in and built upon collaborative tagging systems, in this thesis we explore in detail social tags and their usefulness for enhancing user profiling and resource retrieval. For multimedia data, in particular, such semantically rich labels seem promising as they bring new textual metadata describing the resource.

Second, we analyze friendship links in *Last.fm* contrasting online and "real-world" friends. For this, *Last.fm* offers an interesting ground truth: Users connect to 'online' friends as usual, but they also indirectly reveal their 'real-life' friends by listing events that both physically co-attended. While in *Google+* or *Facebook* users still have to manually maintain circles or lists of close friends, family, etc., we are developing machine learning methods that identify online and off-line friends of different strength automatically. With a special emphasis on the music social tagging platform *Last.fm*, we investigate in depth the following research questions:

- Which kind of tags do users assign to web resources and how frequent are the different types of tags?
- Does tagging and querying behavior correspond, so tags can be used for matching?
- Can we infer additional knowledge, namely the mood or theme of a music piece, by exploiting user generated tags?
- Are socially connected users similar regarding demographics, social network structure, or music preference, and does the tendency correlate with tie strength?
- If so, can we exploit the found assortative patterns as well as transactional information to predict (the strength of) a tie, and which kind of features are most valuable for this task?

Structure of the thesis. We begin by laying the foundations on collaborative tagging systems and online social networks in Chapter 2. To give a first impression on how (much) users engage in both tagging and social networking and by means of which services, we shortly describe a case study in form of a survey among users of the educational bookmarking system CROKODIL[10]. This preliminary background knowledge on systems and their actual usage is then backed up by research conducted in the corresponding research communities. Thus, the related work section (Chapter 3) covers in detail studies on users' tagging behavior, in particular the nature of tags, as well as approaches for exploiting such user generated metadata for enhancing search and recommendation. Focusing on social ties, we also report on scientific results regarding homophily in online social networks and how such user (pair) characteristics can be used amongst others to predict ties and their strength. As with tags, we put emphasis on reviewing methods for leveraging social connections for improving search and recommendation.

The main part of this thesis is then devoted to the presentation of our own research on the value of social tags and social connections in collaborative tagging systems. Chapter 4 describes a thorough analysis of types of tags found in diverse tagging platforms as well as their correspondence with search queries. Based on a self-defined taxonomy of tag types we provide statistics of tag distributions in systems for image, web page, and music annotation as well as distributions of search engine queries for the respective resource types. In addition, we also report on experiments on how to make use of given social annotations to mine additional knowledge. Here, we describe the inference of moods and themes for music tracks, contrasting and combining tag-based methods with algorithms operating on audio content. The following discussion unites the potentials of working with social tags.

In Chapter 5 we look more closely on the explicitly given social connections in tagging systems. We present work on characterizing and automatically identifying online and off-line friendship relations of different strength for the music platform *Last.fm*. Besides analyzing demographics and network topological properties, we are particularly interested in homophily or self-similarity with respect to taste. Our machine learning experiments show that we can reduce the feature set to a handful of features indicative of (strong) friends. Experiments on predicting co-editing behavior on *Wikipedia*[11] show that this task as well benefits from considering social ties.

We close with conclusions and an outlook on issues open for future work.

---

[10] The CROKODIL project is funded by the German Federal Ministry of Education and Research (BMBF) and the European Social Fund of the European Union (ESF).

[11] `http://www.wikipedia.org`

2 Background and Foundations

Both collaborative tagging (also social tagging, social bookmarking, user generated or social annotations) and online social networking are phenomenons associated with the advent of the Web 2.0. The term "Web 2.0" is typically attributed to Tim O'Reilly and the 2004 O'Reilly Web 2.0 conference. It describes a set of new technologies and architectural principles (RSS, AJAX, public APIs, Mashups, and the "Webtop") and, more importantly, a change in how users interact with and on the Web[1]. User behavior changed from passive consumption to active production of content and metadata, and to online interaction. This trend actually made the Web more similar to how Tim Berners-Lee originally supposed the WWW to be – as a Read/Write-Web[2]. Social software supporting communication, collaboration, and the formation of social relationships were on the rise: wikis, bulletin boards, systems for questions & answers, blogging, collaborative tagging systems like *Delicious*, *Flickr*, or *Last.fm* as well as pure social networking services like *Facebook*. The next sections will first lay the foundations for social tagging and then (online) social network analysis. A third section reports shortly on a usage study of social bookmarking and social networking.

2.1 Collaborative Tagging

After a quick introduction on the basic concepts of tagging, we describe the popular tagging systems *Delicious*, *Flickr*, and *Last.fm* – systems studied in this thesis.

2.1.1 Basic Concepts of Tagging

Tags can be described as (key)words or category labels freely selected by users to describe and organize web resources, e.g., for later personal retrieval.

[1] `http://oreilly.com/web2/archive/what-is-web-20.html`

[2] See, e.g., 2005 BBC News interview: `http://news.bbc.co.uk/2/hi/technology/4132752.stm`

> "A tag is simply a word you use to describe a bookmark. Unlike folders, you make up tags when you need them and you can use as many as you like. The result is a better way to organize your bookmarks and a great way to discover interesting things on the Web." [*Delicious*, 2007][3]

Figure 2.1 depicts the most popular tags in *Delicious* (see Section 2.1.2) as of end of 2007. In these so called 'tag clouds', size usually indicates popularity, i.e. usage frequency. Clicking on a tag brings up other resources with this tag. Thus, as the definition states, besides for personal information management, such tags can also be used to navigate (browse) the tagging system to find other resources tagged the same way by other users. Tagging provides a flexible means of information organization as it allows for fuzzy categories and many entry points when browsing, thus making serendipitous encounters more likely. This comes at the price of not having a clear hierarchical structure for navigation. In the inclusive and flat tag set navigation is like keyword based search and by co-occurrences in the network. Though, users might combine general and specific tags to 'imitate' folders or facets.

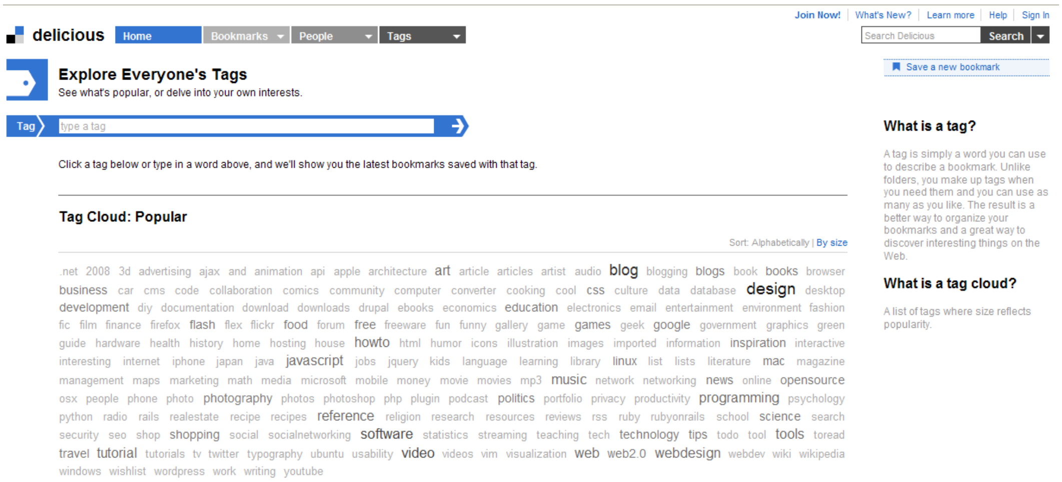

Figure 2.1: Popular tags on *Delicious* (2007)

More formally, a tagging system can be viewed as a tuple of users, tags, and resources (see Figure 2.2). Tags connect users with resources, but they also indirectly link resources and users by tag co-occurrence or tag co-usage respectively. While resources may be directly linked, users may themselves be connected to each other via explicit social relationships. After zooming in into what types of tags we find in different tagging systems, the second part of this thesis will focus on this social aspect of collaborative tagging systems.

[3] *Delicious* was re-designed in 2011. This definition comes from the 2006/2007 version (written Del.icio.us then). Most studies, including ours, refer to this early system before the re-design.

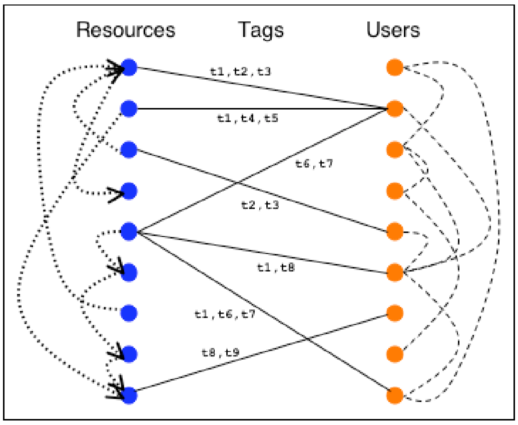

Figure 2.2: A model of collaborative tagging systems (from [155])

Right from when tagging became popular, social annotation and traditional metadata annotation by experts have been often contrasted against each other (see for example [150]). Social tagging has been framed as distributing the workload for metadata creation by building upon "collective intelligence" while top-down approaches on metadata assignment by experts using predefined controlled vocabularies ensure metadata quality. Not controlling tag usage implies spelling mistakes, synonyms (see vocabulary problem as high variability in word usage [81]), or different granularity of tags. Dealing with these natural language problems is not trivial. For example, [50] found that singular and plural forms of a tag may carry different semantics (e.g., observable via different co-occurring tags).

For the global system of tags to be useful beyond re-finding resources tagged by oneself, tagging thus relies on shared cognitive / linguistic structures. In Section 3.1 we will report in more detail on related work showing that tagging systems converge and structure emerges for the popular tags of the folksonomy (folk + taxonomy). The term folksonomy was coined by Thomas Vander Wal as was the distinction between broad and narrow folksonomies[4][157]. "Broad" refers to systems where many users annotate the same set of items. "Narrow" means a resource is tagged by one person or a few, e.g., due to restricted tagging rights. Especially when techniques for feedback are employed, e.g., suggestions for tags frequently used, the border between controlled indexing and free indexing becomes fuzzy [220]. However, there usually exists the "Long Tail" of rather idiosyncratic tags used rarely. Which kind of tags are frequent in a system depends heavily on concrete design choices and, more generally, on a platform's main purpose and users' motivations for tagging [155]. Besides tagging

[4] `http://vanderwal.net/folksonomy.html` and `http://www.vanderwal.net/random/entrysel.php?blog=1635`

rights or tag suggestions, for example, the type of resources to be tagged matters. In Sections 3.1 and 4.1 we will look in detail into tag types in different systems. Quite a variety of systems providing tagging as a central feature exist.

2.1.2 Popular Tagging Systems

The two tagging systems studied the most are *Delicious* and *Flickr*. In 2003, *Delicious* was founded as a social bookmarking system. While browsing the Web, users add bookmarks to their link collection in *Delicious* and assign tags if wanted. Link lists can then be shared with other users – publically or within a smaller user-defined personal network. In one of the first studies on tagging, Golder and Hubermann [90] identified seven common types of tags in *Delicious*, for example, tags identifying what or who some resource is about (e.g., “CSS”, “Rome”), identifying who owns it, identifying qualities or characteristics (“funny”), self reference (“mystuff”), or task organization (“toread”). Note how users tried to overcome the problem of single word tagging in the early *Delicious* system which did not allow for white spaces. Soon *Delicious* provided tag recommendations, i.e., when bookmarking an URL tags were suggested based on tags assigned by other users before[5]. In 2011, AVOS – steered by two founders of *YouTube*[6] – took it over from Yahoo! and re-designed it with the self-claimed focus “on curation and discovery”[7]. There exist a few comparable platforms focusing on managing scientific articles: *CiteULike*[8], *Connotea*[9], and *BibSonomy*[10].

Flickr (Figure 2.3) is a photo-sharing website, where users can upload their pictures, tag and share them – with friends only or the public. The system originated as a by-product of a massive multiplayer online role-playing game. From 2003 on, the stand-alone browser-based platform gained increasing popularity. Since 2005, *Flickr* is owned by Yahoo![11]. In contrast to *Delicious*, users do usually not tag public resources but their own pictures or images by their friends (narrow folksonomy).

Figure 2.4 shows the tag cloud of popular tags currently used in the online social music network *Last.fm*. *Last.fm* was founded in 2002, later merged with Audioscrobbler, and is owned by CBS Interactive since 2007[12]. Since 2009, streaming is partially limited to paying subscribers. As of September 2011 the music community

[5] See, for example, screenshot in `http://www.usfca.edu/uploadedFiles/Destinations/Offices_and_Services/ITS/learning/training/pdf_files/delicious.pdf`
[6] `http://www.youtube.com`
[7] `http://delicious.com/about`
[8] `http://www.citeulike.org`
[9] `http://www.connotea.org`
[10] `http://www.bibsonomy.org`
[11] `http://itc.conversationsnetwork.org/shows/detail1755.html`
[12] `http://techcrunch.com/tag/last-fm`

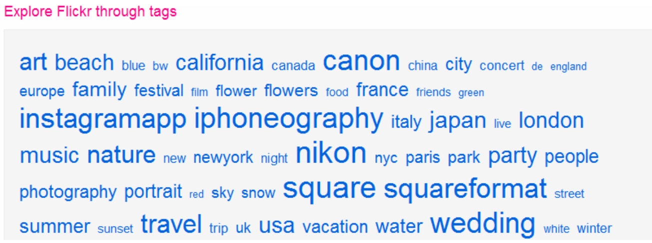

Figure 2.3: Popular *Flickr* tags to be used for exploring pictures within *Flickr*

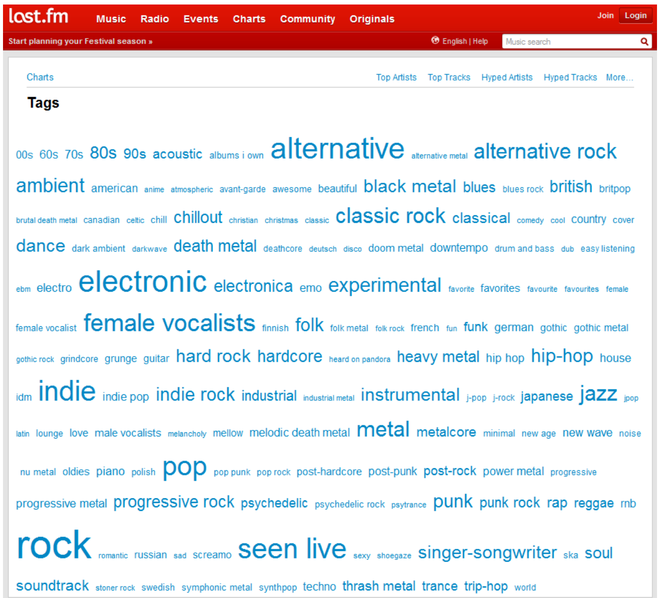

Figure 2.4: Popular tags on *Last.fm*

claims around 40 million users are streaming their personalized radio stations. The core of the system is a music recommender, that can also be installed via plugins to portable devices. With respect to tagging, *Last.fm* can be considered a broad folksonomy, where users annotate the same songs, artists, or albums. Tags can be used to navigate as well as to play tag radio stations, i.e., tracks or artists tagged accordingly. *Last.fm* also supports social features like user walls, friendship networks, groups, mail exchange, an event calendar, and music taste comparison (Tastometer).

2.2 (Online) Social Networks

Employing concepts of graph theory, complex networks like the WWW, the Internet, or the spreading of diseases/epidemics have been studied in a variety of disciplines including physics, biology, linguistics. In the social sciences the focus is on social networks of people having relationships with each other. While early works investigated patterns of interaction in small, closed groups, or ego-networks[13], recent work analyzes huge (online) social networks, e.g., on mobile phone calls [173], instant messaging [134], information propagation through the blogosphere [95], *Facebook*, or *Twitter*[14]. After a short overview of the background on tie strength in social networks, we turn our attention to social network analysis of online social networks.

2.2.1 Social Network Analysis

Network analysis has a long standing research tradition in the social sciences (see [222] for a comprehensive introduction on concepts and methods). Each (social) network can be represented as a graph connecting nodes (or vertices) via edges (links or ties). While in an undirected graph the relation is the same for both nodes involved (e.g., being married), in a directed graph connections are not necessarily reciprocal (e.g., call someone, provide information). Graph theoretic measures, like density, indegrees, outdegrees, centrality, diameter, clustering coefficient, (structural) cohesion, connected components, etc. , describe structural properties and indicate the potentials and bottlenecks of a network. Quite a few software tools (e.g., UCINET[15], Gephi[16], Pajek[17]) have been developed that implement (part of) the aforementioned analytic metrics and provide features for network visualization. For example, a famous finding having received a lot of attention is the 'small-world phenomenon' (also 'six degrees of separation') stating that, on average, every person knows every other person via six persons [212]. Stanley Milgram conducted an experiment [212] where he had 296 volunteers try to deliver mail post to an unknown target by passing it to people they know. For those chains terminating with letters actually arriving, he reported on average six intermediaries connecting sender and addressee.

A plethora of topics exist that are studied in social networks, investigating network structures besides distance. To name just a few (see [222]): centrality and prestige of actors or groups, structural balance and transitivity, cohesive subgroups, affiliations, roles, and positions, etc. Topics and methods vary with the different units of analysis: dyads, triads, larger cliques, or communities. Community detection, for example, is one important area of ongoing research (for a comprehensive review see [76]). A widely used method is the one by Girvan and Newman [89, 171]. It relies on iteratively removing edges with highest betweenness centrality. Edge betweenness

[13] Ego networks consist of an actor (the ego) and the alters, i.e., other actors ego has direct connections with plus all relations between these alters.

[14] `http://www.twitter.com`

[15] `https://sites.google.com/site/ucinetsoftware/home`

[16] `http://gephi.org/`

[17] `http://pajek.imfm.si/`

centrality is defined as the number of all shortest paths connecting any pair of nodes in the network that run along the edge – with traffic or "flow" being equally split across all shortest paths connecting two nodes. The challenge lies in effienctly finding all shortest paths and (re-)calculating betweenness, e.g., based on breadth-first search [67]. Modularity [171] is a popular quality metric for evaluating the goodness of the resulting network split into communities. It achieves high values if many edges lie within communities and only few connect different communities.

One highly influential theory is "the strength of weak ties" hypothesis by Mark Granovetter [92, 93]. From interviews he found that people unusually often received information about job offers via weak ties, i.e. loose acquaintances. His argument then was that such weak ties often connect to other communities – other parts of the global network – and thus function as local bridges. Bridges span structural holes [40, 41] of otherwise not interacting groups, and thus bring the corresponding users the potential benefit of accessing important information (earlier) and to engage in (social) gate-keeping. Strong ties, on the other hand, usually exist within tightly-knit groups, or cliques, which offer trust, mutual support, etc., but they are likely exposed to similar information [92, 40]. The tendency to have ties with people alike is a known phenomenon in social networks (e.g. homophily, see [160, 93]). The stronger the ties, the more probable two people share attributes like age, race, social status, values, or preferences. This bias of often having relations with people similar to oneself can be attributed both to selection mechanisms and social influence. The former implies that people choose to bond with like people partially because we are more likely to encounter similar people in our everyday lives [216, 91]. Social influence results in becoming more similar due to maintaining a connection. The resulting homophily can lead to social circles where diverse opinions and information are missing.

Strong ties also enforce the tendency for triadic closure, a principle saying that two people with a common friend tend to become friends themselves. Reasons for closure may be the opportunity to meet, trust, as well as reducing stress on the side of the common friend. Granovetter stated the strong triadic closure property that the tendency to close the triad (by at least a weak tie) should be even higher if the two existing connections are strong. Otherwise Granovetter calls it a forbidden triad. Using this property it can be argued by contradiction that local bridges are weak ties (see [67], Chapter 3, for the argument). Therefore, weak ties are said to be crucial for innovation, non-redundant information, and network bridging [41, 93]. However, Granovetter's theses remained long time untested on large datasets. One recent study on mobile phone data [173] found that weak ties are crucial for the structural integrity of the network. Strong ties, on the other hand, are important for sustaining local groups/cliques. Concerning information propagation, both types of connections are not sufficient: the first due to infrequent, rare contact, and the latter due to being bound to their local groups.

In general, traditional social network analysis had to cope with experimental design issues, in particular data sparseness. Collecting network data is not trivial for real-world social networks. Common methods include analysis of archival records, observation, and self reports of ego-networks (interviews, diaries) employing elicitation strategies (e.g., "To whom do you turn for advice?")[222]. Besides problems like retrospective informant accuracy, boundary specification is a major issue [222].

2.2.2 Applying Social Network Analysis to Social Media

Online social networking is more and more becoming an integral part of our everyday lives. These networks provide huge and interesting datasets to revisit earlier findings and sociological theories, thereby overcoming certain design issues of early 'real-world' social network studies. Most platforms support public user profiles, which besides basic demographic attributes also provide information about taste preferences (e.g. bookmarks), the friendship network of a user, as well as observable interactions with others (wall posts, commenting). For many platforms, such data can be collected (given user consent) via publically available APIs. Thus, large network samples can be gathered.

Social interaction on the Web may also deviate from findings for the off-line world. Thus, the analysis of online social networks can identify emerging trends. Recent work, for example, found that on *Facebook* (as of early 2012) average distance has shrunk to around 4.74 on average, i.e., 3.74 or four degrees of separation [11].

Most popular is the general purpose platform *Facebook*, which starting of in 2004 is meanwhile used by over 900 million users (as of April 2012[18]), and earns its money via targeted advertising. It outran *Myspace*[19], which is nowadays pretty much focused on music. On *Facebook* users have their wall, i.e. user profile, which shows basic profile information, uploaded pictures, pages or entries rated positively ('Likes') or commented on, and posts from their friends. Friendship is undirected, formed after mutual consent on confirming a friend request. Users can interact with each other via wall posts, chat, private messages, groups, or simply by liking or commenting on each others activities, e.g., from the news feed showing activities of one's friends. Part of

[18]`http://www.pcmag.com/article2/0,2817,2403410,00.asp`
[19]`http://www.myspace.com`

Facebook's success was its openness. It allows developers to build and integrate small applications (many of which are games like Farmville or Poker). Recently, *Facebook* provides JavaScript based "Social Plugins" like the Like-Button or friend activity feeds to be integrated within private or company websites. It is not only for such collection of user data external to *Facebook* that privacy concerns are raised again and again.

Google started in 2011 with *Google+*, introducing directed, non reciprocal relationships and social circles as the main distinctive feature. Social circles are user defined arbitrary groupings of friends that can be used to channel information sharing and to filter the news feed. The micro-blogging platform *Twitter* is a further global player regarding online social networking. Its key features are "tweets", short messages of maximum 140 characters. People can follow certain users and spread of tweets can be restricted to one's followers. Tweets can also be directed to specified users (@username) and they can be grouped by using hashtags introduced by an '#'. Besides, there exist a variety of smaller platforms popular in certain areas or countries as well as special purpose online networks, for example, *XING*[20] or *LinkedIn*[21] both targeting business professionals.

On such community sites people connect with their close 'off-line' friends, with loose acquaintances rarely met, or they make friends with unknown, like-minded people. Motivations for connecting are manifold: staying in touch, socializing, finding experts, exchanging knowledge and information, etc. As a consequence, the explicit, often binary friendship relation between two people does not reflect well the true 'hidden' relation and its strength. For example, as of early 2012 a *Facebook* user has on average 190 friends [11].

Marlow and his colleagues [43] showed for *Facebook* data of 2009 how explicit friendship links and 'real' interaction differ. They contrasted four types of networks defined by different notions of friendship ties: the explicit *Facebook* friendship links, links representing reciprocated communication, one-way communication links, and maintained relationships, i.e., ties established by clicking on a news feed story of a person or by visiting her/his profile twice or more. As is depicted in Figure 2.5, the authors found that though users tend to have large friend lists the number of people they keep track of is considerably smaller (around 50). The set of friends they actually interact with is even smaller (10 to 20). Hence, the number of people we really maintain relationships with is expected to be a lot smaller than the average *Facebook* friend list. Based on results from primates Dunbar argues that humans

[20] `http://www.xing.com`
[21] `http://www.linkedin.com/`

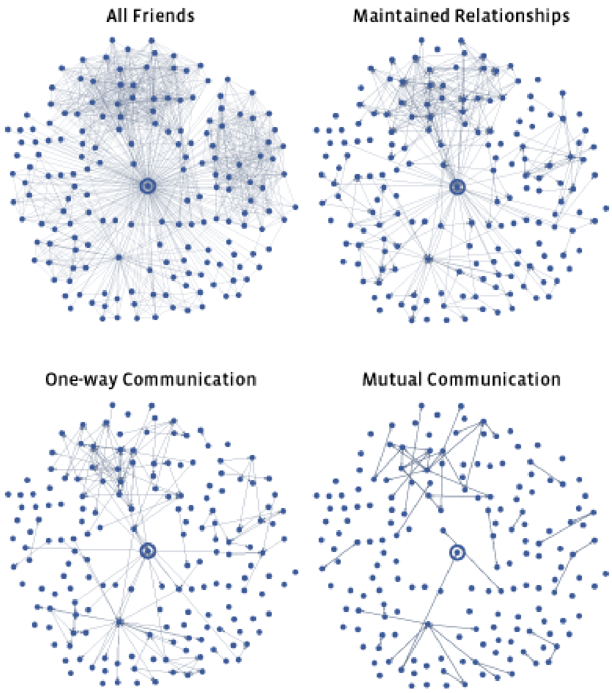

Figure 2.5: Four kinds of ties users have with their *Facebook* friends (from [43])

can cognitively handle around 150 (the so-called Dunbar's number) relationships with a very restricted set of close ones [65]. Similarly, [224] compare the network formed by explicit social connections with the actual interaction graph for *Facebook*. Again, interaction is limited to a few friends only. Consequently, both graphs show considerable differences when it comes to standard network measures like diameter, average path length, etc.

In their analysis on 309,740 users of the micro-blogging platform *Twitter*, Huberman et al. [113] found that posting behavior is driven by a sparse hidden network of 'actual' friends, a subnetwork of the declared set of followers and friends. Attention being a scarce resource, the ratio of those real friends a user directly posts to compared to all followees is very small with an average of 0.13. Though the authors' friendship definition is directed, this keeping in touch interaction pattern is reciprocated for around 90% of a user's friends. However, *Twitter* is not highly representative for online social networks as directed one-to-many mass communication, especially by elite users like big media organizations, celebrities, etc. to their followers, constitutes big part of communication on the social platform [226].

Explicit binary links, thus, do not indicate actual interaction, closeness, or attention. Popular sites like *Facebook* or *Google+* reacted recently by introducing friend lists and social circles respectively. However, both have to be manually created and maintained. *Facebook* owner Mark Zuckerberg is quoted to admit: "But guess what? Nobody wants to make lists"[22]. The work presented in Chapter 5 aims at automatically inferring strong and weak ties to be used for (semi-)automatic generation of lists like "close friends" or "acquaintances". Next, we shortly report on a case study on which and how collaborative tagging systems and online social networks are used.

2.3 A Case Study: Usage of Tagging and Participation in Social Networks

In order to get a better picture of user behavior regarding bookmarking and tagging as well as their participation within social communities and Web 2.0 tools, we shortly report on results of a paper-based survey (preliminary results were published in [6]) conducted among people in (re-)education[23]. The participants are part of the target groups of the CROKODIL project, which develops a platform for better supporting Collaborative Resource-Based Learning, that is, collaborative learning based on web resources. It follows a quick overview on the project's goals and motivations as well as the current version of the platform.

Given the rapid growth of online knowledge bases and, at the same time, the pace at which knowledge is outdated, using resources found on the Web for knowledge acquisition is becoming increasingly important. Such resources may be open learning content from educational institutions (like iTunes U[24]), user-generated content like on *YouTube* or *Slideshare*[25] as well as collaboratively constructed resources such as wikis and blogs [102]. Besides using and generating content, learners collaborate with other learners using different applications like social networks, discussion boards, wikis, or forums. However, such self-directed collaborative learning based on web resources imposes a lot of challenges on the learner. This comprises, amongst others, the phrasing of search terms, selection of relevant and trustworthy web pages (e.g., from results returned by a search engine) as well as organization and structuring for later use [189]. The overall goal in CROKODIL is to support all these tasks in one platform by developing and combining novel pedagogical and technological concepts.

Within the platform (see Figure 2.6) users can store and annotate resources found on the Web[26], they can associate it with learning tasks or goals (so-called "activities"),

[22]`http://techcrunch.com/2010/08/26/facebook-friend-lists/`

[23]Items in the questionnaire were phrased in German and are translated here for comprehension.

[24]`http://www.apple.com/education/itunes-u`

[25]`http://www.slideshare.net`

[26]Storing and tagging resources while searching the Web can also be done via a Firefox plugin.

share them with co-learners, and search for other('s) resources. Potentially relevant resources on the platform are also actively recommended to the learner based on content or metadata as well as usage similarity and social connectivity. With respect to organizing resources via tags, CROKODIL supports semantic tagging by allowing users to add a concept type like "person", "location", "topic", "event", "type" to each tag. For encouraging and supporting the collaborative use of these resources and their information, social networking functionalities are provided in the platform; they are complemented by integrating external social networking services (*Facebook*). In order to enable use in more formal instructional settings, traditional learning management systems can be connected via widgets.

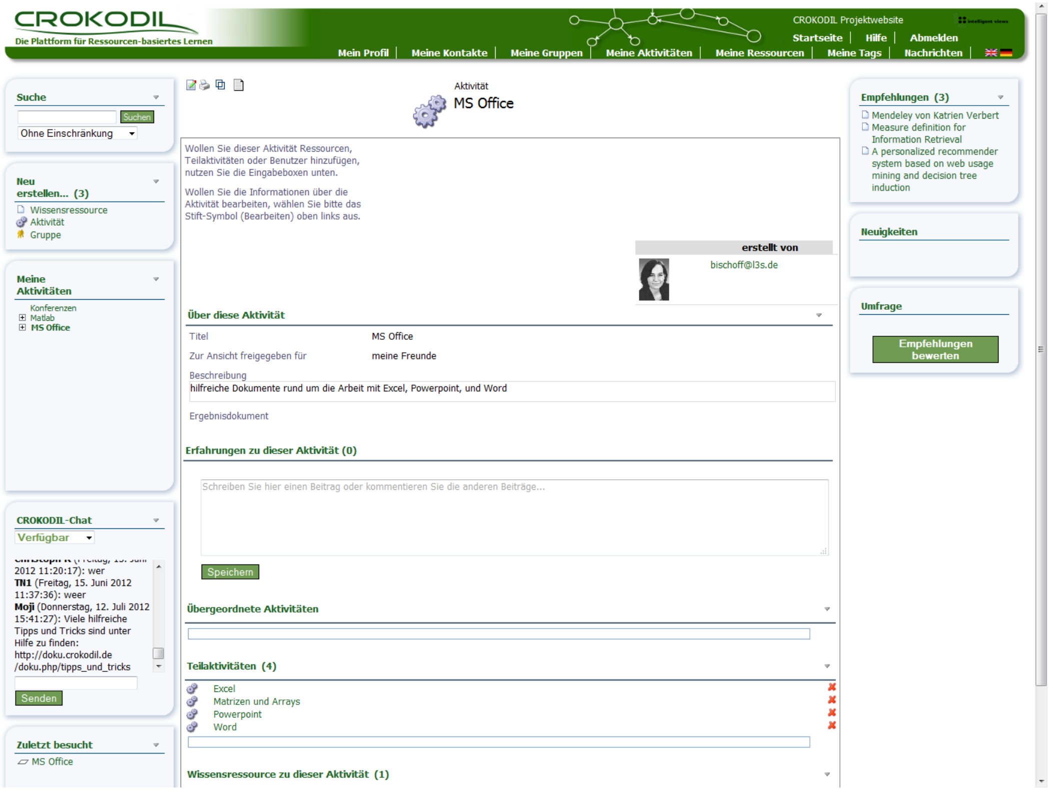

Figure 2.6: Screenshot of the current CROKODIL prototype

Coming back to our survey on CROKODIL target users, the participants are enrolled in the following educational programs:

- Bachelor of Arts in Business Administration + Industrial Clerk / Bachelor of Engineering + Electronics Technician for Automation Technology (Group 1)
- education program for school dropouts (Group 2)
- re-training in information technology (Group 3)

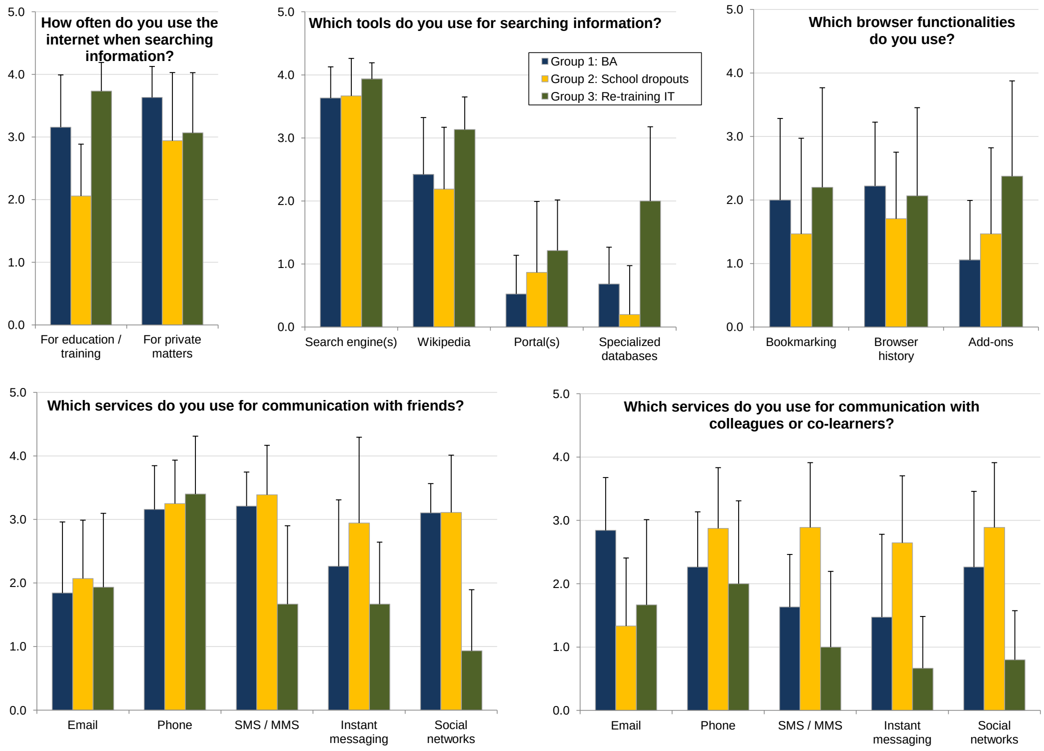

Figure 2.7: Selected results of a usage study on bookmarking and social networking

From Group 1, we evaluated 19 participants: 12 male, five female, two not specified, with an average age of 19.6 and a standard deviation (SD) of 0.96. Group 2 has 11 female and seven male participants, and it is also quite homogeneous regarding age, having an average age of 18.7 (SD 1.6). Of course, the age difference is considerably higher for the older participants in re-training (average 32.8, SD 7). Here, male participants predominate (14 male, one female). Most of the survey items were rated using a five-point Likert scale ranging from zero (never) to four (almost always). Figure 2.7 shows selected results. Error bars indicate the standard deviation from average values.

All participants of the survey are intense Web users, going online once or several times a day and feeling quite confident handling this medium (averages between 3.3 and 3.7). Exceptions were only found among the school dropouts: three participants go online “a few times per month” ‘only’, four “more than three times per week”. All three groups predominantly use the Internet at home. At work it is also used oftentimes by the BA students of Group 1 and the learners in IT re-training. The latter as well oftentimes use the Internet during re-training activities (avg. 3.4, SD 0.63) as the primary means for searching information with respect to re-training tasks (avg. 3.7, SD 0.46). For these learners in Group 3, the most popular Internet activities are private surfing, email, online shopping, and education related activities. For the students in Group 1 and the school dropouts of Group 2, communication services like email, instant messaging, and social networks are a lot more prominent.

The participants from Group 3 show a heavy reliance on search engines and the *Wikipedia* when browsing the Web to find relevant information. Portals and specialized databases are sometimes visited too. Some participants also mentioned community related bulletin boards or newsgroups as valuable sources, they sometimes refer to. The participants in Groups 1 and 2 almost exclusively rely on search engines and to some extent the *Wikipedia* as well. Expert groups or specialized databases are not used at all. With respect to organizing and re-finding relevant web resources, tools like bookmarking and browser history are used infrequently. Only some users regularly make use of bookmarking and the history. Similarly, tagging is not a very common practice in neither group. Half of the users in Groups 1 and 3 do not know the term tagging and few infrequently tag themselves pictures, videos, or web resources. In Group 2 only 30% are familiar with tagging. Those who are, however, use tagging frequently, in particular for pictures and videos.

In our evaluation, *YouTube*, *Facebook*, and its German variants *MeinVZ / StudiVZ / SchülerVZ*[27] are the most popular social networks. We found some significant differences between the groups with respect to social networking. Only the younger users of Group 1 and 2 (around 2/3) actively use these platforms on a regular basis. Here, *Facebook* and its German counterpart *MeinVZ / SchülerVZ* are used by almost all respondents; more than half of them (68% and 55%) frequently visit these sites to communicate with friends, acquaintances, and co-learners, to publish content, to stay up-to-date, and to give or recieve recommendations.

Social networks are also important for communication with both friends and co-learners or colleagues. For private communication, they are almost as popular as phone or SMS contact. In professional communication, only email is used more often. For the re-trainees in Group 3, in contrast, communication with friends or colleagues/co-learners within such social platforms is still rare compared to more traditional media like telephone or email. Preference of 'face-to-face' communication was also mentioned explicitly a few times. This difference in the usage of social networks can be most probably explained by considering the age difference between the groups. *YouTube* serves as a source for finding interesting content for the participants of all three groups.

[27] `http://www.studivz.net`, `http://www.meinvz.net`, `http://www.schuelervz.net`

3 Related Work

Now that the general background for collaborative tagging systems and the social ties within them has been set, this chapter reviews in detail related work done in the respective research communities. We start by describing users' tagging behavior and the resulting characteristics of social tags. We then review approaches for supporting search and knowledge management that were developed based on such findings. The second part of the chapter delves into prior work on online social networking, in particular on the study of homophily and the prediction of ties and their strength.

3.1 Collaborative Tagging

Research on collaborative tagging has mainly run along two lines: studies of user behavior, including user motivations as well as patterns in tagging, and studies of how to make use of the newly available (meta)data to improve search and recommendation. After summarizing the most relevant work in these areas, a third subsection points out how this thesis advances the state of the art.

3.1.1 Tagging Behavior

Golder and Huberman [90] were one of the first to analyze usage patterns in *Delicious*. Amongst others, the authors find that after around 100 bookmarks the proportions for each tag per resource remain stable – with more general tags having higher proportions. The stabilization process can be modeled via a stochastic urn model, and it is attributed to imitation and shared knowledge. Similarly, Halpin et al. [101] found a scale-free power law that forms within a few months (see also [157]), indicating consensus on top tags plus a 'long tail' of less frequent or even idiosyncratic tags. The authors present a generative model based on shuffling, which results in tag convergence or stabilization in form of a power law distribution attributed to the phenomenon of preferential attachment, also known as the "rich get richer" effect. In the shuffling model an existing tag is re-used with a constant probability P(o), i.e., its relative usage frequency so far, and a new tag may be introduced with a probability of 1-P(o). Of course, in reality tags are not introduced at random, but the authors name information value considerations as the cause. The latter highlights the trade-off between tagging and search using tags. While according to the principle of least efforts speakers prefer ambiguous, general terms, hearers prefer words with high entropy not to have to post complex or/and multiple queries [101]. The folksonomy structure emerges at the intersection: consensus though tagging is mostly for personal benefit [101].

Based on their analysis of 58,728 posts to 64 URLs, Kipp and Campbell [124] came to a similar conclusion: Consensus on a small set of highly popular tags is accompanied by inconsistencies that do follow several predictable patterns. [44] present as well a model building upon preferential attachment, but the underlying stochastic model is modified such that selection of an existing tag is done based on a power law kernel, capturing recency and frequency effects as in human memory.

In the study by Golder and Huberman on *Delicious* [90], seven types of tags are identified differing in the functions they fulfill: identify what (or who) it is about, identifying what it is, identifying who owns it, refining categories (like numbers), identifying qualities or characteristics, self reference, and task organizing. Marlow et al. [155] helped formalizing the analysis of collaborative tagging systems by proposing a taxonomy of dimensions in system design, pointing out potential implications the different design choices may have. For example, tagging rights (self-tagging like in *YouTube*, free-for-all as in *Last.fm*) as well as object types (textual, video, pictures) and source of the material (user-contributed, global) influence the type and nature of tags used within the system. Tagging support, e.g. tag suggestions, can impact how well and fast the folksonomy converges.

An experimental study on the *MovieLens* movie recommender system[1] [191] shows how the tagging community, i.e., conformity and social proof, influence convergence in the folksonomy through visibility of community tags as a form of tagging support. Users of the system were randomly assigned to one of four groups: no display of other peoples' tags, display of all tags, display of popular community tags, and display of automatically recommended tags. In the latter two conditions tagging behavior showed stronger convergence, namely on factual tags (82% and 67% resp.). Interestingly, in the shared condition (display all tags) 60% of tags were classified as subjective while in the private scenario 39% were personal and 38% factual.

[1] `http://movielens.umn.edu`

Other dimensions of the tagging system taxonomy by Marlow et al. [155] are: aggregation model, resource connectivity and social connectivity of users. The authors also name possible user motivations and incentives, some mainly 'organizational' in nature, some purely 'social': future retrieval, contribution and sharing, attract attention, play and competition, self presentation, and finally opinion expression.

In an exploratory study on *Last.fm* and *Amazon* [239], opinion expression, performance including self-presentation ("Crime against humanity", "make it stop", "seen live") as well as activism ("defectivebydesign") were found emerging social motivations characteristic for such free-for-all, broad tagging systems. Often, such tags have a lot of characters and are phrases. For self-tagging of personal resources in *Flickr* (combined with the mobile application *ZoneTag*), in contrast, self organizational (retrieval, directory, search), social organizational (contribution, attention, ad hoc photo pooling), self communication (context, memory), and social communication (content descriptors, social signaling) are identified as motivations [4].

With the goal of suggesting tags for web pages based on existing social tags, the authors of [229] propose criteria for good (sets of) tags: high coverage of multiple facets, high popularity, least-effort, and uniformity. Certain types of tags (e.g. organizational tags) are to be excluded. For this, a simple tag taxonomy consisting of content-based tags, context-based tags, attribute tags, subjective tags, and organizational tags is presented. Sen et al. [190] use implicit user behavior data and explicit user ratings on tags for predicting high quality tags in *MovieLens*. System data on the number of users having assigned or searched for a tag are good indicators of tag quality. The total number of tag usage may be misleading, though, as it potentially gives high weight to a few power users. Displaying valuable tags should then influence other users to also tag 'better' by re-inforcing such quality tags. In [192], the feature set is considerable extended: the number of times a tag is assigned to a specific item and the average number of times a tag is applied to the item it annotates are the single best performing features.

In [104], a hierarchical category model for tags is created based on tags in *Connotea* and is used to describe tag usage within the system. On a first level, linguistic category model, functional category model, and tag to text category model are differentiated. Word class, spelling, neologisms, and language are subcategories in the linguistic category model. The majority of *Connotea* tags are single word nouns. Within the functional category model, subject related tags (92% of the 1,191 analyzed *Connotea* tags for 500 articles) can be resource or content related while non-subject related, personal tags are either affective, time and task related, or fall under tag avoidance (no tag). The category 'tag to text' deals with redundancy or novelty value of tags from an information retrieval point of view: tags are identical to fulltext words including title and abstract (54%), they are variations of it (16%), or they do not appear in the fulltext (30%). Comparing tags and author provided keywords, it was found that *Connotea* users assigned less, simpler, and broader terms.

In their comparative analysis Heckner et al. [105] use this taxonomy to manually classify around 1000 tags each from *YouTube*, *Connotea*, *Flickr*, and *Delicious*. More than 91% of the tags in all systems are subject related, except for *Flickr* (76%). Most of those subject related tags (89% and more) are content-related. From the resource-related tags, creator is frequent in *Connotea* and, though considerably less, also in *YouTube*, device is popular in *Flickr*, and resource type in *Delicious*. Almost no affective, or time and task related tags are found. The few affective tags found are usually positive or neutral, and their majority comes from *YouTube*. Regarding informational value, tags typically do not appear in the title of the resource (66% for *YouTube* and *Delicious*, 73% *Connotea* and 86% *Flickr*). An interesting phenomenon of "overtagging" is observed for *YouTube*, where users provide many, sometimes synonymous tags probably to ensure findability for other people.

Comparing motivations for personal information management vs. resource sharing in *Delicious*, *Flickr*, *YouTube*, and *Connotea* in detail, Heckner et al. [103] conducted a user study with 142 participants. They found a significant difference between *YouTube* and *Delicious*. In *YouTube* users upload and tag with the goal of sharing content with others while in *Delicious* users tag for personal re-finding. Qualitative evidence is reported for a focus on sharing in *Flickr* and personal information management in *Connotea*. With respect to sharing, *Flickr* users share pictures with friends and family while in *YouTube* friends and family, colleagues and neighbors as well as unknown others are targeted likewise.

3.1.2 Exploiting Tags for Search and Recommendation

In this section we will provide details on methods aiming at exploiting tags for improving search and recommendation. After a quick overview on general approaches, we will zoom in on how to enhance music retrieval via tag-based knowledge mining.

3.1.2.1 General Strategies for Searching and Recommending with Tags

Research on search and recommendations in social tagging systems is manifold. Focusing on navigational aspects, for example, Dubinko et al. [64] automatically identify interesting tags and visualize their evolution over time. In [9], user tags are combined with content-based techniques for improving data navigation and search for images. Yet, other work aims at improving information access by extracting explicit semantics from tags in folksonomies, e.g., to induce tag hierarchies by extracting ontologies from folksonomies (see e.g., [236, 188, 37]). Related is the prediction of certain tag types, e.g., events and places using burst analysis on geotagged and timestamped *Flickr* pictures [179] or into 'Location', 'Artifact/object', 'Person/group', 'Action/event', 'Time', 'Other' using the lexical database WordNet[2] [199] combined with *Wikipedia* [174]. With the goal of overcoming problems of free tagging, SemKey [153] as a system supports semiautomatic typed tagging. The users select tags, disambiguate their meaning based on a list of concepts, and they specify the relation to the resource ('hasAsTopic', 'hasAsKind', 'myOpinionIs').

Several approaches making use of tags within recommender systems have been proposed[3]. Recommender systems aim at suggesting users additional, new resources that correspond to a user's interest and, thus, should be relevant for her/him. Such systems are based on explicit and implicit (inferred) user preferences, assuming that items similar to what a user liked before will be considered relevant. Two main approaches to estimating similarity can be differentiated.

Content-based recommender (see, e.g. [177]) analyze resource contents or metadata (keywords, tags) to find related items matching the user profile, i.e., extracted keywords or tags from documents the user likes. Similarity between profile (query) and document terms is often computed using standard Information Retrieval metrics like cosine similarity (see, e.g. [183]). Collaborative-Filtering (CF) algorithms, in contrast, ignore resource contents. Instead they rely on user interaction with resources. In the case of user-based Collaborative Filtering, the first step is to find users with similar interests to then recommend items positively rated by those neighbors that are unknown to the user (see [184] for a short introduction). Those neighbors are calculated based on the overlap of similarly rated items (e.g. Jaccard similarity). In item-based Collaborative Filtering, on the other hand, item-to-item similarity is estimated first based on co-usage or co-rating by users (e.g., *Amazon* and their "Costumers who bought this item also bought y"). Problems with CF arise as the approach heavily depends on a large amount of user ratings or other feedback data in general as well as for the single user ('cold-start problem'). Second, novelty and diversity are often missing.

Applying Collaborative Filtering for recommending resources, people, or tags in folksonomies is not straight forward, though, as traditional Collaborative Filtering

[2] `http://wordnet.princeton.edu`

[3] For an overview on approaches for recommender systems in social tagging systems see [154]

works on two-dimensional matrices and not on ternary relationships, i.e. three dimensional arrays [154]. For example, tags have been a means to extend the user-item matrix in Collaborative Filtering by (adding) a user-tag matrix in case of user-based Collaborative Filtering and by (adding) a tag-item matrix in case of item-based Collaborative Filtering [213]. A fusion of both approaches performs best in recommendation experiments on *Last.fm*. In [234], exploiting tags during (implicit) query expansion with similar tags improves coverage to also include relevant, but less popular resources from the 'long tail'. Second, ranking recommendation candidates partially based on tagging similarity between the user and the users associated with the corresponding resource is beneficial for accuracy.

Nakamoto et al. [169] propose to incorporate tags into recommender systems by comparing the tag vectors of bookmarked or rated resources to capture the context of such (implicit) interest indication when (a) computing user similarities and when (b) predicting item ratings. In later work [170], tag vectors were replaced by topic domain vectors derived from the resources' tags in order to overcome problems like synonymy or morphological variants. Also, a user's current interest – modeled as the domain vector of the website viewed at the moment – is considered when choosing from candidate resources.

In [74], tags have been proven useful for recommendation of music tracks. While Collaborative Filtering over the user-tag matrix performed worse than a standard track-based Collaborative Filtering baseline, searching through tagged songs with user tag profiles directly performed best – in particular when the tag-profiles are derived indirectly by taking all those tags from *Last.fm* that were assigned to the songs a user listened to. Interestingly, users seem to be rather conservative taste wise: There was a high inverse correlation between the participants' ratings of track preference and novelty. Sen et al. introduce "tagommenders" [193], systems that recommend resources (here movies in *MovieLens*) based on preferences inferred for tags, not on the items themselves. Signals for tag preference can be tag application, tag search, or tag quality rating as well as indirect signals like clicking or rating the items annotated with a tag. In the recommendation experiments tagommenders outperformed traditional Collaborative Filtering in generating recommendations, but they were not superior in predicting exact user ratings for resources.

Recommendation of additional (personalized) tags has been a highly active research area as well (see, e.g. [116, 128, 127, 205, 199, 168, 42, 229, 162, 204, 37]). Suggesting appropriate tags for a resource is said to reduce the cognitive burden of tagging as it changes the task from generation to re-cognition [204].

Similar to our work, [106] studies the value of tags and URLs of bookmarks in *Delicious* for improving web search. As positive aspects the authors state that, amongst others, most tags are judged accurate by humans, and that there is a significant overlap between popular tags and popular queries in the AOL log [176]. However, the value is somehow limited as specific tags are often associated to certain domains/hosts (e.g. 'java'). More importantly, 50% of the tags are already in the page's content and 16% in its title. Similar results have been reported in [141]: Around 50% of the URLs have all their associated tags already in the full text.

Experimenting with the predictability of social tags based on page information like text, anchor tags, and surrounding hosts (links to the page), [107] found that actually many tags are redundant as they can be predicted with around 90% precision. However, this only holds as long as a small recall (here 10%) is acceptable. Regarding recall, the authors also found that association rules can increase recall for queries with a sinlge tag, probably capturing synonyms or polysemy.

[231] propose to exploit social bookmarks by combining them in form of SBRank (number of bookmarks for a page) with the link-based approach of PageRank (the core algorithm of Google [175]). For the around 1,300 URLs analyzed correlation of SBRank and PageRank was 0.53, indicating partial overlap of result sets returned by both metrics as well as a potential for complementing each other. With SBRank, the authors argue, page popularity would not be defined by page authors only, and the dynamics in bookmarking may compensate for PageRank penalizing young/new pages. Moreover, sentiment tags can provide a useful proxy for quality assessment. An algorithm for enhanced search using a linear combination together with re-ranking, e.g., of sentiment aspects, is qualitatively evaluated.

Bao et al. [15] suggest to incorporate social tags into search via SocialSimRank and SocialPageRank. While SocialSimRank identifies similarities between tags by iterating over the bipartite graph of annotations and pages (with user counts as edge weights), SocialPageRank iteratively propagates popularity from pages to users to annotations back to web pages back to annotations back to users back again to pages. Comparison of SocialPageRank and PageRank regarding annotation popularity proves that SocialPageRank gives high values to pages that are popular among taggers. Experiments show that considering one or better both simultaneously improves retrieval performance considerably compared to a standard baseline with and without PageRank. Similarly taking into account the triadic structure (users, resources, tags) of folksonomies, [111] proposed Adapted PageRank and the topic-sensitive adaptation "FolkRank" considering preferences.

Zhou et al. [238] present a generative probabilistic model for web documents and their annotations, namely an adapted version of Latent Dirichlet Allocation (LDA) [32] accounting for tags as well. Retrieval experiments prove the value of tags in addition to document terms, especially in form of tag-based topics for users or documents. In [39], a very similar generative model was extended to also include the user, who first decides to bookmark a resource based on her/his topical interests. Tag prediction based on the new hierarchical model outperforms traditional LDA, and personalized tag recommendation is superior to baselines exploiting tag popularity.

Investigating the graph theoretic network metrics of characteristic path length and clustering coefficient, Cattuto et al. [45] found small world properties in *Delicious* and *BibSonomy* as well as distinctive connectivity characteristics of spam tags given a weighted network of tag co-occurrences. Co-occurrence patterns have also been exploited to discover topics and communities of interests, e.g., in [141], where topics are represented through frequent tag sets, mined by association rules with topics being then used to cluster users and URLs together. Evaluation on *Delicious* data showed that the average pairwise cosine similarity within a cluster is considerably higher than between clusters. Such topics cover most of the individual user's tags and are thus appropriate for user profile modeling. In [37], however, single social tags were less suited for clustering together similar blog posts than is Google News' "related"-mechanisms for clustering its news articles.

In the next section we will focus on enhanced search for music through knowledge enrichment, i.e., automatic classification based on social tags.

3.1.2.2 Enhanced Music Retrieval via Tag-based Knowledge Mining

Music enrichment recently focuses on deriving mood information based on extracted acoustic data [72, 147, 202]. For example, [147] proposes a content-based method, tailored to classical music, that uses music features on intensity, timbre, and rhythm for classification along Thayer's model on human emotions [209] (see Section 4.2.2.1). In a related work [72], relative tempo, mean and standard deviation of average silence ratio are used to classify (using a neural network) the mood of a track as corresponding to "happiness", "sadness", "anger", or "fear". However, such approaches are not able to capture aspects external to the music piece's content such as social context or usage context, associations people have with music they know.

To overcome these shortcomings, several existing papers aim at automatically inferring additional information from available content as well as (user generated) metadata. Turnbull et al. [214] present a music retrieval system that uses supervised multiclass Naïve Bayes classification for learning a relationship between acoustic features and words from expert reviews on songs, thus enabling query-by-text for music.

Similarly, [18, 215] aim at enriching songs with textual descriptions for improving music retrieval. [18] uses a variant of the AdaBoost algorithm, FilterBoost, in order to predict social tags of the songs based on the information captured in the audio features. There is no special focus on mood and theme-related tags, like in our case, but the tags learned by the classifier pertain to multiple categories of tags (genres, styles, moods, and contexts). [215] compares five methods for collecting tags: user surveys, harvesting social tags, annotation games, mining web documents, and auto-tagging audio content. Again, there is no discussion about the performance of the described methods for predicting mood and theme tags. Moreover, both are not comparable with our approach since there is no clear definition for mood and theme classes, and the data sets on which evaluation was performed differ from ours.

In [68] as well as [206], user tags are explored for improving music recommendations. While the first work focuses on mitigating the cold-start problem by predicting additional tags based on acoustic features and their relationship with existing tags, [206] aims at better recommendations based on the latent factors hidden in user-tag-item relations. For this, the authors successfully apply higher-order singular value decomposition on the triplets.

A combination of social tags and content-based features has been successfully used to automatically predict music genre [48]. Two strategies are proposed that make implicit use of *Last.fm* tags: a graph of music tracks is constructed that captures their semantic similarity in terms of tags associated. Both the baseline low-level feature only classifier as well as a single-layer classifier, considering audio features and implicit tag similarity simultaneously, are clearly outperformed by a double-layer classifier which firsts learns genre labels based on audio information, and then iteratively updates its models considering the tag-based neighborhood of tracks.

For non-textual multimedia data collaborative annotations are particularly valuable, bridging the semantic gap by adding information that low-level features do not capture. User tags in *Last.fm* (the semantics of) span a vector space of low dimensionality covering sensible attributes as well as similarity of music [137]. Retrieval experiments show that tags are well-suited for enabling automatic organization of tracks by genre or artist. The authors also analyze mood-related tags: A self-organizing map trained on tags corresponding to emotion words exhibits similarities with the often used two dimensional models of valence and arousal.

In [139], social tags are fused with audio-based "muswords", generated from timbre and rhythm, of automatically identified regions of interest in music tracks. Evaluation measures mean average precision on retrieval experiments querying with a track for other tracks with the same genre or artist. The results show that content-based muswords do not provide a significant benefit as long as tracks are tagged with at least two tags. With fewer tags, however, inclusion of audio is helpful. The integrated approach of muswords and social tags still outperforms state-of-the-art content-based methods when a third of the songs has only a single tag, which is highly likely an appropriate genre label [139]. Training semantic aspect models (Probabilisitic Latent Semantic Indexing, PLSA) first on social tags and then learning the latent topic associations for muswords helps overcoming the segregation of sparsely-tagged tracks and tracks tagged a lot. In prior work [138], the authors used similar retrieval experiments on social annotations finding that PLSA is superior to its non probabilistic variant LSA – so is the simple vector space model.

Somewhat complementary to our approach, [112] aims at studying the relationships between moods, artists, genres, and usage metadata. As a test set for the experiments the authors use *AllMusic*[4], *Epinions*[5], and a subset of *Last.fm* data. The authors find that the relationship between mood and usage is especially unstable because of the specific terms and phrases used to denote the usage activities. The authors point out an interesting finding: Many of the individual mood terms were highly synonymous, or described aspects of the same underlying mood space. The experiments also showed that decreasing the mood vocabulary size in some ways clarified the underlying mood of the items being described.

3.1.3 Contributions of the Thesis

Our work in Chapter 4 extends prior work analyzing tagging systems by establishing a simple, but comprehensive tag classification scheme that is applicable to different systems with different kinds of resources. In contrast to most earlier work, we investigate questions regarding user tagging behavior contrasting diverse popular tagging systems: *Delicious* for general web pages, *Flickr* for pictures, and *Last.fm* for music resources. The study on *Delicious*, *YouTube*, *Flickr*, and *Connotea* published in [105] is very close to our work, establishing a complex hierarchical taxonomy and manually classifying tags accordingly. Yet, we report descriptive statistics of tag type prevalence not only for each system in general but also by categorizing very popular,

[4] `http://www.allmusic.com`; *AllMusic* has an expert maintained classification system for music, e.g., along genres, moods, and themes (see Section 4.2).

[5] `http://www.epinions.com`

somewhat frequent as well as highly idiosyncratic tags. Most importantly, we discuss the value of user generated tags based on the found patterns in tagging. For this, we contrast tags with queries from a search engine log – unlike [106] assessing different types of queries for different types of resources and of varying popularity.

Regarding knowledge mining based on tags, we add experiments for predicting mood and theme labels for music, i.e. *Last.fm*. While earlier approaches explore the use of *Last.fm* [68, 206, 215], audio content [18, 215], Web documents, surveys, or annotation games [215] to predict (the likelihood) of all kinds of tags, our work explicitly focuses on inferring (the underrepresented) mood and theme annotations. In particular, we complement work on automatically classifying music mood/emotion based on audio features (e.g., [147, 72, 202]) by using *Last.fm*'s valuable folksonomy information for inferring mood as well as theme labels for songs. Whereas in earlier experiments only tags were used for deriving moods, themes, and styles/genres [24, 26], here we also investigate fusion with audio-based methods. Extending existing music metadata enrichment studies, we fuse social tags and low-level audio features of the tracks to infer mood and theme labels, showing that both sources provide helpful complementary information.

3.2 Social Ties and Tie Strength

An impressive amount of work has been done in the social sciences regarding the analysis of self-similarity along social ties (see [160]) as well as the measurement of tie strength (see, e.g. [156]). For the latter, a lot of reliable indicators have been identified, e.g., interaction frequency, duration, intimacy (e.g. [156]), network topology (e.g. [40]), mutual friends (e.g. [198]), social distance (e.g. [146]), recency (e.g. [145]), reciprocity (e.g. [80]), etc. ([87] for a quick overview). Here, we shortly review work on online social networks.

3.2.1 Homophily in Online Social Networks

Some recent studies started addressing the topic of homophily in online social networks to provide first large-scale support for the hypothesis. Analyzing a Microsoft Messenger instant messaging (IM) network, [134] found a strong tendency of preferred communication with self-similar others, especially for language, location – with the frequency of conversation decreasing with increasing distance –, and age. For gender, the opposite was true. For *Facebook*, [14] also confirmed the inverse relationship between distance and friendship likeliness. Given user provided address data, the author find a power law with an exponent near -1. This distribution is then used in a maximum likelihood approach to predict the location of users knowing the address of one or more friends.

Studying balance of social attention in *Facebook*, Backstrom et al. [10] observed gender homophily in messaging especially with females. Such within-gender communication is more dispersed than across-gender communication, which usually concentrates on a few top friends of the opposite gender. For online dating, [73] found significantly higher interaction between men and women matching with respect to marital status, number of children wanted, physical build, and smoking habits. In [38], friend behavior was highly predictive of what a user votes on in a platform for voting political statements though similarity in attitudes was not necessarily high. In a family feud online game played in pairs, people preferred to interact with people similar in terms of gender and education [55].

Association by similarity has also been found in a university online network for most user self-descriptions as well as for interests – the more 'niche' or 'social' the hobby the higher the tendency [3]. The authors report moderate support for the weak link hypothesis: Dissimilar people are important as cross-community links/bridges. Mislove et al. [164] analyzed similarity among *Facebook* friends. They found strong affinity with respect to attributes like department, college, or high school. However, for the more complex attribute 'political views' they only found a very weak tendency for assortative mixing.

Studying tags and groups as implicit preference attributes, Schifanella et al. [187] found a clear trend of online friends behaving similar in *Last.fm* and *Flickr*. For active users in *Last.fm*, number of friends, frequency of tagging as well as semantic similarity correlate among friends beyond random. For the tagging platform *CiteULike*, [132] report higher self-similarity between connected users on interests measured via shared items, metadata, and tags. In both studies similarity decreases with increasing distance in the social network.

Groh and Ehmig [94] found a medium correlation between friendship pairs' ratings of clubs in Munich, with the tendency becoming stronger when focusing on cliques of three or four. For their scenario, (groups of) friends are thus more similar in ratings than groups of people who do not know each other. In their social network analysis of *Orkut*[6], *LiveJournal*[7], *Flickr*, and *YouTube*, Mislove et al. [163] also reported high reciprocity and strong positive correlation on link indegrees of connected users. Yamamoto and Matsumura [230] analyzed optimal heterophily between senders and

[6] `http://www.orkut.com`
[7] `http://www.livejournal.com`

receivers in terms of blogging influence (tracked via re-occurring terms and links) and domain knowledge. They found that the majority of pairs favor small heterophily. In particular, people most often adopt topics or products when the sender is just slightly more influential.

For *Twitter*, Macskassy and Michelson [151] showed that retweeting of posts (tweets) can be partially explained by user profile similarity – independent of how well the current tweet fits the interest profile. Here, user profiles and tweets are modeled as topics corresponding to *Wikipedia* categories. Though similarity was in general low for most user pairs, similarity was considerably higher than expected, and the homophily model better fits observed local information propagation behavior than the three other models investigated: a general model based on tweet recency, a model accounting for communication recency, and a topic model considering similarity between the user profile and the tweet at hand. However, for many users, a combination of models best explains retweeting.

In [226], homophily of attention could be shown for four classes of "elite" *Twitter* users: Celebrities follow and retweet to celebrities, media to media, organizations to organizations, and bloggers to bloggers. For classifying users into "elite" and "ordinary" the authors leveraged *Twitter* lists, i.e., user-defined groupings of other users, thus tapping the potential of the "wisdom of the crowds".

An analysis on *Myspace* [210], a platform centering around music and videos, gave evidence of homophily with respect to attributes like ethnicity, religion, age, country, and marital status. There was no sign of homophily for gender. Focusing instead on explicit statements about music, books, videos, etc. on user profiles in *Myspace*, Liu [148] found that, interestingly, the 29,979 users' tastes were less similar to the averaged taste of their top eight friends than to 'common sense' tastes on *Myspace*.

Baym and Ledbetter [16] surveyed 559 active *Last.fm* users asking them to characterize the relationship to one random friend, e.g., how much shared taste and musical history motivated the formation of their *Last.fm* friendship. The authors apply hierarchical multiple regression on user rated friendship strength. Besides high similarity on age, users reported to share musical taste with their friends. However, taste is not predictive of tie strength, so does not help differentiating weak from strong ties. Communication across all platforms was best indicative. In general, tie strength – rated on a five-point Likert scale – was rather weak below the scale midpoint. The authors suggest that shared taste fosters online friendship, which rarely turn into strong relationships. However, a majority of users stated that the relationship had began outside of *Last.fm*.

Interestingly, though online social networks offer a means to communicate and collaborate more anonymously and to overcome time and geographic boundaries, it might actually increase the tendency of bonding with people alike by facilitating organization into communities of self-similar people [182].

[57] investigate the interplay between similarity and social ties, trying to isolate similarity among friends as due to selection or due to the process of social influence, i.e., becoming more similar to conform with neighbors. The authors present a networked urn model considering social interactions as well as edit profiles (personal, from friends) to simulate communication and editing behavior observed in *Wikipedia*. Clear feedback effects between both variables are found. The value of the social interaction network vs. a similarity-based user network for predicting future behavior seems to depend heavily on the concrete social system under consideration. Likewise, for *Facebook* data taken from two different points in time, [129] show that both factors impact similarity: social influence and homophily.

With the similar goal of distinguishing the role of peer influence and homophily in network contagion with respect to adoption of a mobile service application, [7] employ a matched sampling procedure comparing adoption behavior of nodes with adopter friends to nodes likely to have – but actually not having – the same number of adopter friends due to similarities in user characteristics and behavior. Thus controlling for homophily as an explanation for co-adoption, the experiments show that traditional approaches comparing to shuffled, random networks overestimate the impact of social influence up to 700%, especially at early stages of the contagion process. Homophily explains more than 50% of the observed contagion and partially the higher adoption rate given clusters of adopters. Tie strength modelled as messaging volume (instant messaging) influences adoption behaviour, even controlling for homophily.

3.2.2 Tie and Tie Strength Prediction

Regarding the automatic prediction of social ties in online networks, two main tasks can be differentiated: predicting the existence of a tie between two people and classifying a known social link with respect to its strength or its type.

3.2.2.1 Predicting The Existence of a Tie

As one of the first, Adamic and Adar [2] predict social connections based on Web data like homepage text, inlinks, outlinks, and mailing list membership. Comparing various measures proposed for link (strength) prediction (e.g., Adamic and Adar [2], Katz measure [120], preferential attachment, etc.), Liben-Nowell and Kleinberg

[142] show for a co-authorship network that considering network topology alone already leads to substantially better link prediction performance than random guessing. Newer work [53] applied link prediction based on such structural information to the blogging platform *LiveJournal*. Given dynamic data, the authors find that most connections are established within 10 days after joining the network. [143] make the case for framing link prediction as a supervised learning task, thus overcoming domain-specificity and data imbalance issues associated with unsupervised topological metrics or approaches like [2, 142]. In [13], a supervised random walk algorithm combines structural information with user attributes for link prediction on *Facebook*.

Analyzing mobile phone call data, Eagle et al. [66] show that friendship reveals itself through characteristic temporal and spatial patterns of co-occurrence. Due to cultural norms reciprocal friendships load on the extra-role factor, i.e., time spent in close proximity after campus hours or during the weekend and outside of campus. The authors also find that satisfaction within one's working group is correlated with having friends in proximity while calling friends from work indicates exactly the opposite. Comparing to self-reported relationships as well, there is overlap with the observed mobile data. However, there are differences, e.g., due to recency effects.

Similarly, [58] examine how the number of co-occurrences of two people in time and space point towards an existing social tie. Applying a proposed probabilistic model, which accounts as well for homophily regarding the likelihood of picking the same place to visit, to geotagged and timestamped *Flickr* data, the authors can well generate the friendship probabilities observed empirically: Friendship is massively more likely than baseline if two people have at least a few co-occurrences within short time ranges and small distances.

[109] predict follow back behavior on *Twitter*. Their graph model is based on the finding that two-way relationships are structurally balanced, that is, in triads of people either all three relations are present or only one. Also, the authors report homophily to be present in reciprocal relationships for status and time zone as well as for social connections. In order to better understand network evolution (in *Twitter*), [161] propose a method for inferring from a single snapshot when links were formed.

Relying on semantic similarities of user tags and associated items in *Last.fm*, Schifanella et al. [187] predict binary online friendship better than the platform's own recommender. For the *Yahoo! Pulse*[8] network, indirect interest connections via user applications/services and direct friendship links are mutually helpful, e.g., for predicting friendship [232].

[8] `http://www.pulse.yahoo.com`

A related area is the recommendation of new people to connect to in order to extend one's personal network (e.g., [97, 47]). While the goal of establishing new social ties is quite different from predicting actually existing ties and their strength, similar data is used for the task: demographic data, group membership, co-tagging, and structural network information. For example, [97] compare different similarity sources from tagging, bookmarking, blogs, forums, and friending. They find that overlap with familiarity is rather small and that recommendations based on the different sources produces diverse sets of people. Golder and Yardi [91] showed that the wish to form new ties in *Twitter* is influenced heavily by transitivity and mutuality of the existing social network.

In a user study on the IBM enterprise social networking site Beehive [47] it was found that relationship-based approaches like Friend-of-Friend (FoF) generate significantly more recommendations of known contacts, resulting more often in connecting. As content-based methods produce more good recommendations of unknown people, an intelligent fusion of both approaches should be aimed at. Interestingly, it seems that known people are usually rated as good recommendations though that not necessarily means users also want to connect. Too many social contacts and too weak ties were reasons mentioned by the participants.

3.2.2.2 Predicting The Type or Strength of a Tie

Again related, but different in nature, is the prediction of types of ties, e.g., "family", "colleague", etc., for given unlabeled relationships. [208] present a factor graph model exploiting user pair attributes, correlations between relationships, and global constraints in a semi-supervised learning task optimizing parameters for fitting the partially labeled network data. In [207], an extended approach for classifying relationship types across different partially labeled social networks is presented. For this, the model incorporates the social network theories of social balance, social status, opinion leadership, and structural holes to complement features specific for a certain (type of) network. Evaluation of this knowledge transfer, e.g., between *Epinions* and *Slashdot*[9], or between a co-author (ArnetMiner[10]) and the Enron email[11] network, shows the value of such additional information.

Similar in spirit, [63] identify relationships of the type manager-subordinate on the Enron email corpus. Given labeled ground truth data, a ranking function is learned operating on varied features. Here, content-based ranking outperforms ranking based

[9] `http://slashdot.org`
[10] `http://arnetminer.org`
[11] `http://www.cs.cmu.edu/~enron`

on email traffic. Top terms for the relationship considered are: "please", "report", "project", "termination", and "executed". [221] propose a time-constraint probabilistic factor graph model maximizing likelihoods for mining advisor-advisee relationships from a collaboration network of co-authors of articles in computer science.

[135] predict positive and negative links on *Epinions*, *Slashdot*, and *Wikipedia*. For example, in *Slashdot* users can name "friends" and "foes". On *Epinions* trust and distrust can be specified. Building upon balance and status theory, degree features (positive and negative indegree, positive and negative outdegree, total indegree and total outdegree, common connections) are used together with 16 triadic features (one feature for each type of possible triad involving the link under consideration) for training coefficients in a logistic regression model to classify the sentiment of given links. The approach achieves high accuracy of around 80% or higher, depending on the dataset and the embeddedness of the edge. The authors also show that knowing about negative links, not only positive ones, is beneficial for predicting the presence or absence of a positive link.

Kahanda and Neville [117] recently presented a machine learning approach to automatically identify strong friends. The authors formulated a link strength prediction task: For each friend pair (u,v), given their user profile attributes like age, gender, etc., their interactions (writing on the friend's wall, tagging a photo), and network information (e.g., number of mutual friends) a supervised learning method decides whether they are "top friends". Evaluation on data from the public Purdue *Facebook* network, where users can nominate best friends within the "Top friends" application, showed that best friends can be successfully distinguished from weak ties. The best classification results were achieved by using bagged decision trees on network-transactional features (i.e., moderate transactional activity like wall posts by interactions with other users), which account for 97% of the performance observed using all features. Thus, user interactions are highly predictive, especially when viewed in context of user behavior within the larger social network.

In a similar work, Gilbert and Karahalios [87] predict tie strength as a linear combination of 74 *Facebook* variables (e.g., last comment, number of friends, wall words). Predicting tie strength on a continuous zero to one scale the authors achieve a comparable classification performance of 85%. Intimacy (number of friends, initimacy words), intensity (wall words, outbound posts, thread depth), and duration (first comment) are most indicative of tie strength. While the specific predicitive variables and weights may not move beyond *Facebook*, the provided mapping of different variables to (sociological) dimensions should allow for generalization.

In later work [86], the authors show that the model with its weights as trained on *Facebook* largely generalizes to *Twitter*. For this, the top predictive variables are mapped to their (closest) *Twitter* equivalents. 2,114 people made use of the "We Meddle" *Twitter* application which creates lists of inner and outer circles, i.e., strong and weak ties, to be used for filtering *Twitter* feeds. Quantitative evaluation builds upon corrections made by users deleting wrongly placed persons from the list(s). With an upper bound of 15.7% error rate the model's performance is comparable to the quality of its predictions within *Facebook*. It also shows similar shortcomings: It can not deal with intense negative relationships nor with changes in tie strength through time. With respect to individual predictors, network structural features differed most, indicating that it is difficult to map directly from *Facebook*'s explicit, binary friendship network to *Twitter*'s follower network.

Also focusing on *Facebook*, [194] present a model of tie strength calculation based on a weighted addition of online interactions (wall posts, comments, messages), face-to-face interactions (photo tags), and interest-based interactions (groups, shared events). The model takes time into account (via gradual forgetting) as well as relationship relevance (defined as the number of persons engaged in interaction/members of a common group). The authors report anecdotal evidence for three users on how tie strength varies depending on the parameters chosen for the proposed formulas.

In [96], an axiomatic approach of inferring tie strength in implicit social networks is presented. It is based on a bi-partite graph of persons and the events (real events but as well email exchange) they participated in. Similar to the work in [142], the authors analyze popular measures from prior work with respect to how well they satisfy the set of eight axioms for tie strength measures (e.g., frequency: the more interaction the stronger the tie). The authors do not find one particular best function, but many measures are appropriate regarding these axioms.

[228] propose a latent variable model assuming the strength of a relationship to be the concealed effect of user profile similarity and the underlying cause of user interactions. Thus, given profile similarities, relationship strength is modeled as conditional probability drawn from a Gaussian distribution. The model is evaluated by using the derived tie strength to identify *LinkedIn* user pairs having same attribute values. Tie strength outperforms prediction based on, e.g., similarity or interaction count. For *Facebook* data, tie strength outperforms other graphs (e.g. friendship) in increasing autocorrelation, i.e., the concordance of attribute values on connected nodes, on gender and relationship, less so for political and religious views. It also performs best in a collective classification task of predicting the same four attributes.

To differentiate personal from professional closeness, Wu et al. [225] look at relationship multiplexity on a workplace social networking site, performing regression analysis on interaction data. Evaluating against a ground truth of 196 users' ratings of their relationships, the model achieves around 71-82% accuracy. Yeung and Iwata [8] focus on trust ties established on product review sites like *Epinions* and *Ciao*[12]. Over time strong trust ties tend to differ less in their ratings. While a trust relation does not ensure preference similarity, the authors' extended matrix factorization approach leads to tie strength estimates that correlate with user similarity.

3.2.3 Exploiting Ties for Search and Recommendation

Approaches for efficiently searching and propagating information in online communities build strongly upon methods developed in social network analysis. Epidemic or gossip-based algorithms adopt patterns established in (communication) networks to enable efficient spread of information for distributed computing, or to request or query routing in Peer-2-Peer systems (see, e.g., [62, 61]). In order to cope with malicious attacks, trust is a particularly important topic in such peer networks [52, 118] though the notion is not the same as in social networks.

Similarly, social search and recommendation algorithms try to exploit the communication and interaction patterns found in social networks as well as, e.g., the trust and similarity typical of strong ties. Referral Web [121] is a first approach to integrate social networks and Collaborative Filtering. A social network was constructed from names co-occurring in the WWW, e.g., links on a home page or co-authorship. Queries that can be answered based on this network have the form "which connection do I have to XY" or "documents about databases by people close to XY".

A newer system is Aardvark [110], a social search engine build upon the village metaphor of asking people questions looking for highly contextualized answers or subjective recommendations. In Aardvark, user questions are routed to potential answerers based on topic expertise, availability as well as social connectedness. The latter is defined as a combination of mutual friends and affiliations, demographic and profile similarity, and behavioral characteristics such as verbosity, vocabulary, speed, chattiness, and politeness match. First evaluation results show the feasibility of the paradigm for certain kinds of information needs. Social proximity is important as answers by people closer in the social network were given better quality ratings.

SmallBlue [69], similarly, is a search system for finding experts and communities within the intranet of a large company. Here, expertise topics and social connections

[12] `http://www.ciao.co.uk`

are extracted from private mail. For a given query, social distance is then displayed for each person in the result list. SONAR [98] provides a framework for such applications, with interfaces for storing and collecting information on social connections collected from private or public data, which also incorporates tie strength between people. SONARBuddies is one example client, in which contacts are ordered by connection strength. In [119], a social relevance score considers, besides textual relevance, how much the information seeker trusts another user as well as how intense the latter user interacted with a recommendation candidate.

Tie strength may also be exploited in the design of enhanced communication interfaces like ContactMap [223], which organizes communication around people within a social desktop to better emulate social workplace functions like social reminding (communication commitments, keeping in touch) and social data mining (social recommendation, tracking project status). While users were not satisfied with a fully automatic identification of important contacts based on email communication frequency and reciprocity, a later prototype employed similar extraction tools to aid manual construction of contact maps.

[203] models real-world information flows in order to give recommendations and rank users according to influence based on the usage of certain communication paths. For this, diffusion rate between users is computed based on access time/order to the same documents. Automatic evaluation shows that standard Collaborative Filtering algorithms can be outperformed in accuracy by up to 80%. Moreover, the underlying social network can be used to overcome data sparsity, e.g., by applying factor analysis on the user-item-matrix enriched with explicit user connections [149].

Chen and Fong [49] present a framework on how to make use of social relationships within a Collaborative Filtering recommender system. Here, relationship strength is composed of a similarity component (accounting for overlap in demographics, interests, activities like groups or events, etc. and applications) as well as a trust aspect. For the latter, the authors estimate the relative importance of the single trust factors, namely the groups of *Facebook* variables as used in [87], by employing a C4.5 decision tree on a user survey on *Facebook*.

[122] present a framework for recommending new people in social networks, considering amongst others tie strength when computing pairwise user similarities in generating candidate lists. For example, the tie strengths to all of a user's friends, modeled as the weighted sum of system dependent communication and interaction features like number of emails sent, should hint towards the general social openness of the user.

For personalized recommendations of new posts concerning a news item, [195] extend their Collaborative Filtering recommender system such that strong social ties (here: members of a thematic group) indicate a high value of a post with respect to completeness and simplicity. Weak ties, in contrast, imply diversity of opinions. From ratings given to posts, the system learns a user's preference regarding completeness and diversity, to which recommendations are adapted. [186] presents a framework for social search and recommendation that integrates classical Collaborative Filtering attributes for users and resources with an ontology and social connectivity (explicit friendship or 'spiritual', i.e., similar interests modeled via tags) within a scoring model. A small evaluation study shows that 'spiritual' connections in particular improve search results significantly – but not for all kinds of queries. Social query expansion by tags used by friends, however, did not lead to improved performance.

In a related work, [17] demonstrated that social search, implemented as search among all friends having used a query term as tag before, possibly combined with an authority score for users can yield the best precision for search in *Flickr*. Also for efficiently searching inside collaborative tagging networks like *Delicious*, incorporating social connections between users and between tags proved useful. A top-k algorithm combined with dynamic tag expansion and dynamically extending search over socially connected users can answer queries considerably faster than traditional approaches [185].

Some work, e.g., on collective classification, is making use of information of connected or similar users to predict private attributes of users [164, 237]. For example, Zheleva and Getoor [237] predict private attributes of *Flickr* and *Facebook* users based on friendship links and group membership information. Assuming half of the sensitive attributes to be known from public user profiles and exploiting friendship links in the adjacency matrix, they are able to predict with 56.5% accuracy the location of a *Flickr* user and with 68.6% the gender of a *Facebook* user. While friendship links only worked for certain attributes, group information achieved even better results. In a related work, Mislove et al. [164] generate communities based on shared attribute values to predict missing values for friends. For some attributes, given around 20% actual values only, this approach achieves over 80% accuracy.

3.2.4 Contributions of the Thesis

Our research on the music network *Last.fm* and the online encyclopedia *Wikipedia* complements studies on real-world networks as well as the recent research on tie strength in general purpose online networks like *Facebook* and *Twitter*. Thereby, we transfer the problem of tie (strength) prediction to the taste domain, where preference or interest similarity may be an important ingredient, and is thus to be considered in much more detail.

While many earlier studies required users to manually rate tie strength directly (e.g., [16, 117, 225, 87]) or indirectly (e.g., "How would you feel asking this friend to loan you $100 or more?" [87]), *Last.fm* offers an interesting proxy for tie strength: physical co-attendance at events listed in the event calendar. In particular, we add experiments contrasting such off-line ties with online ties, as two notions of friendship, on a rich set of factors extracted from the digital records people leave on *Last.fm* – a type of platform rarely studied.

We extend prior work by Baym and Ledbetter on *Last.fm* [16], the work that is closest to our study, by analyzing observable behavior within the platform instead of considering questionnaire like data, i.e., explicit user ratings of friendship strength, musical taste or reported interaction with friends. Similarly, actual user behavior may not correspond (completely) with self-declared statements about interests regarding music, books, videos, etc. on a user's profile as studied in [210] on *Myspace*. Also, *Myspace* was in its early years less focused on music, but it was more a general purpose social network. Here, we exploit a variety of implicit user preference indicators like tracks listened to, favorite artists, tags used, etc. Thus, in contrast to link prediction in *Last.fm* based on user tags [187], tagging semantics are not in the focus of our methods for automatic tie prediction (Chapter 5). We will consider it as one indicator out of many for predicting online and, in addition, off-line friendship links.

As a further experiment, we predict future co-editing behavior in *Wikipedia* based on social networking data. While in [57] the probability of co-edits was based on either common friends or article similarity, we augment this work by taking a variety of typical tie strength indicators like communication frequency or recency as well as network metrics like clustering coefficient, betweenness centrality, etc. into account.

4 Absent and Potential Ties: The Value of User Generated Tags

As they offer a promising way to estimate similarity between resources, users, and resources, or between different users, the usefulness and reliability of tags is important for many search and recommendation algorithms. Even in absence of explicit social connections, the sheer amount of the semantically meaningful social tags hold the potential to show patterns arising through so called 'collective intelligence'. These can help to identify recent trends in topics or, in the simplest case, to retrieve resources based on the newly available metadata, provided by a variety of users describing the resource and its content, e.g., regarding its topic (e.g. "databases"), type (e.g. "video"), associated time or locations or more subjectively indicating preference and opinions (e.g. "funny"), or associated contexts and usage (e.g. "vacation"). For example, as one famous 'game with a purpose' [219], the ESP game[1] supports the collection of textual metadata for images in order to improve retrievability. Similarly, tags used by a user can facilitate elaborated user modeling, enabling representation of and matching according to a user's topical interests.

To prove that tags are indeed a useful source of additional information, in the first section the focus will be on analyzing tag usage patterns and their implications for user profiling, search, and recommendation. In particular, we will study the kinds of tags used, their frequencies in different tagging systems, and we will compare them to search engine queries. Building upon the found characteristics, we will then present approaches exploiting tags to enrich resources or user profiles with additional information: music moods (e.g., "mellow", "energetic", "angry") and themes (e.g., "party time", "chill", "wedding songs"). Both kinds of labels are valuable as general music perception – i.e., how we think and talk about music – is heavily influenced by emotions and context. Searching for music usually is an exploratory and social

[1] `http://www.gwap.com/gwap/gamesPreview/espgame/`

process, in which people make use of collective knowledge as well as the opinions and recommendations of other people [133]. Related is their need for contextual metadata expressing, for example, which situations/events are often associated with the songs. Thus, besides directly searching or browsing music by artist or title, associated usage, theme/main subject, and mood/emotional state are used in every third (navigational) query [133]. As our analysis below will show, though often queried for, themes in particular are underrepresented in the music tagging system *Last.fm*.

4.1 The Usefulness of Tags for Profiling and Search

Web 2.0 tagging platforms like *Delicious*, *Flickr*, and *Last.fm* have made online information organization and sharing through tags popular, mostly because it is so easy. First studies, in particular on the first two sites, indicate that tagging motivations and as a consequence the resulting nature of tags differ across systems based on, e.g., resource type, tagging rights, connectivity, etc. [155]. As not all tags are equally useful for user profiling and search, this section studies in detail tags found in different tagging systems. Besides establishing a tag type taxonomy suitable for multiple domains, we compare tag type distributions for tags of varying popularity, and as well contrast them with types of user queries posted to search engines.

4.1.1 Datasets

In order to be able to answer the questions raised not only for one specific tagging system but for systems with varying features and content, we examine tagging behavior in *Last.fm*, *Flickr*, and *Delicious*.

Last.fm. During May 2007 we collected a large set of data by crawling *Last.fm* pages for tags, tracks, and users. This way, we ended up with about 317,058 music tracks and the corresponding metadata like title and artist as well as user provided tags with their usage frequencies. Using the most popular tags as a starting point, we gathered about 21,177 distinct tags assigned by users to tracks, albums, or artists. Besides total usage frequencies, the number of distinct users and related tags, including a score of similarity, were stored for every tag.

Flickr. The *Flickr* data we studied was kindly shared with us by the University of Koblenz/Landau and the Tagora project. Starting from the beginning of 2004 until the end of 2005, some of the most popular tags were used as seed tags for subsequent expansion of the crawl. The 100,000 pictures crawled first and the accompanying 32,378 distinct tags are used for our analysis.

Delicious. The data analyzed here was made available to us by the Knowledge and Data Engineering Group/Bibsonomy at the University of Kassel. During four days in July 2005 an initial set of around 6,900 users and 700 tags were collected from the start page of *Delicious*. Recursively, further data was gathered using these seed items. Further users as well as resources were gathered by monitoring the start page. In total, a few thousand usernames were used to retrieve the 10,000 web resources tagged first by every user. From this data the relevant information was extracted: resource URLs, associated tags, time, descriptions, usernames. As a result, we ground the following analysis on a set of 323,294 unique tags assigned to 2,507,688 bookmarks.

The distributions of usage frequencies follow approximately a power law curve in all three systems, reflected by a straight line on a log-log plot. For *Flickr* and *Last.fm*, however, we find a sudden drop, which is probably a consequence of crawling based on popular tags. Looking closer at the slopes for each system (ignoring absolute frequencies), we observe that tag frequencies are rather evenly distributed in *Flickr*. For *Delicious*, influence of popular tags is a little more pronounced. *Last.fm*, in contrast, lies at the other extreme having the steepest slope: with a few very popular tags and 60% of the top 100 tags representing genre information. This is not surprising, since music as a specific domain has a narrower vocabulary than pictures or web pages, which can show whatever person or object or treat a multitude of topics.

4.1.2 A Tag Type Taxonomy

To better understand what kinds of tags users assign to the different kinds of resources, we first need to define a taxonomy of tag types. This categorization scheme should capture all distinctions that appear relevant when tagging a resource, whether this regards the content or metadata, idiosyncratic organizational cues, etc. After a review on existing tag classification systems [90, 191, 229], we adapted the one by Golder and Huberman [90] to include Time and Location as well. This refinement is important when generalizing the scheme to be also applicable to systems like *Flickr* or *Last.fm*, having different resource types. As we will see, for pictures, for example, the latter two categories are frequent while they are negligible when it comes to web pages in general.

The scheme was improved multiple times based on estimating interrater agreement on a subset classified according to each taxonomy version. Table 4.1 shows the resulting final classification scheme with examples found in the three systems. For this final classification system, we had a good and substantial inter-rater reliability of κ 0.71. As the standard measure to assess concordance for our nominal data, Cohen's Kappa (κ) [51] indicates the achieved interrater agreement beyond-chance. It is noteworthy that classification consistency was higher for the more focused, narrower systems *Last.fm* and *Flickr*. When allowing the raters to name a second category in case of ambiguity, κ was boosted to 0.8 (0.9 for *Flickr*).

Category	***Last.fm***	***Flickr***	***Delicious***
Topic	love, revolution	people, flowers	webdesign, linux
Time	80s, baroque	2005, july	daily, current
Location	england, african	toronto, kingscross	slovakia, newcastle
Type	pop, acoustic	portrait, 50mm	movies, mp3, blogs
Author/Owner	the beatles, wax trax	wright	wired, alanmoore
Opinions/Qualities	great lyrics, yum	scary, bright	annoying, funny
Usage context	workout, study	vacation, honeymoon	review.later, travelling
Self reference	albums i own, seen live	me, 100views	wishlist, mymemo

Table 4.1: Tagging taxonomy with examples for *Last.fm*, *Flickr*, and *Delicious*

Topic describes the content of the associated resource, i.e., what it is about. This may be the theme of a song reflected in the lyrics, or a person or an object depicted on a photograph. For textual web resources, such topical information can often be extracted from the site's content [106]. *Time* tags give information about the temporal context of an item, e.g., when a photo was taken, a song was produced, or a website was created. For the spatial dimension, *Location* captures the place a picture was taken or a musician/band originated from. Some tags indicate the *Type* of media or the file format of the resource tagged (e.g., "mp3", "weblog"). For music, here we also include instruments and genre, specifying the type of music. With photography, in contrast, type rather refers to the (artistic) style of a picture (e.g., "portrait", "lomo"). *Author/Owner* names the creator or owner (e.g. record labels) of a resource.

Tags may also express subjective *Opinions*, for example, for mostly socially motivated self-presentation in free-for-all-tagging systems or alike. On the other hand, such subjective statements may be used like a rating for future retrieval by oneself (e.g. "funny") or to share it with others as a recommendation. Tags falling into the *Usage context* category give hints on what a resource could be used for or on the context/task it belongs to (e.g., "travelling", "jobsearch") – though partially subjective, these are still potentially useful even for unknown users. As the last class, *Self-reference* tags are truly personal in nature, referring to the person tagging. As such, these tags will not be beneficial for enabling enhanced search of resources or for mapping interests of users based on their tag profile, e.g., for social search or tie formation purposes. However, the information contained can nevertheless provide relevant information to another user of the system, e.g., giving a clearer picture of a person, who, for example, has been recommended as potentially interesting contact.

In [22] we showed that the different types of tags proposed in our taxonomy can be automatically identified. Five of the eight tag classes were determined with the help of look-ups in lists/knowledge bases and rules expressed in regular expressions: Time (regular expression, pre-defined list), Location (geographical dictionary, so called gazetteer, from GATE[2]), Type (resource dependent lists including music genres from *AllMusic*, file and media formats, photographic styles), Author/Owner (artist database for music, regular expressions for websites), and Self-reference (list). Due to their openness and dynamics, i.e., their multitude, using predefined lists or knowledge bases is not reasonable for the much broader categories Topic, Usage context, and Opinion.

Instead, we employed machine learning via the C4.5 decision tree [178] implemented in the machine learning library WEKA [99][3]. The C4.5 algorithm builds a decision tree for classifying data by recursively splitting the training data along the attribute with the highest normalized Information Gain. Information Gain ranks attributes based on how much information is gained when considering that attribute for predicting the class, i.e., how many bits are saved for encoding the class value knowing the attribute's value. Underlying is Shannon's information theoretical concept of Entropy [196, 152], measuring the bits necessary to encode or transmit a variable's value, thus indicating a variable's purity or uniformity of distribution and thus the uncertainty associated with predicting its values. We trained on the following features: distinct users of a tag or usage frequency (for *Flickr*), number of words and characters, part of speech, and semantic category provided by the lexical database WordNet. For evaluating our methods we compared against a manually labeled set of tags, partially used in the tagging behavior analysis described.

[2] `http://gate.ac.uk`
[3] `http://www.cs.waikato.ac.nz/ml`

	Topic	Time	Location	Type	Author/ Owner	Opinions/ Qualities	Usage context	Self-reference
Delicious	**67.00**	**1.00**	**3.86**	**8.00**	**6.29**	**5.14**	**7.86**	**0.86**
top 300	65.33	0.33	3.33	15.00	3.00	3.33	8.67	1.00
2nd 200	73.00	1.50	4.00	4.00	5.00	6.50	6.00	0.00
3rd 200	63.50	1.50	4.50	1.50	12.50	6.50	8.50	1.50
Flickr	**46.00**	**4.86**	**26.14**	**5.29**	**0.14**	**7.43**	**7.57**	**2.57**
top 300	35.67	8.67	34.00	5.33	0.00	5.33	9.00	2.00
2nd 200	47.00	1.50	20.50	7.50	0.00	11.50	7.50	4.50
3rd 200	60.50	2.50	20.00	3.00	0.50	6.50	5.50	1.50
Last.fm	**2.43**	**1.29**	**8.29**	**51.14**	**8.14**	**17.71**	**6.43**	**4.57**
top 300	1.33	2.33	10.33	62.33	0.33	12.33	5.33	5.67
2nd 200	1.00	1.00	10.00	51.00	6.50	19.00	6.00	5.50
3rd 200	5.50	0.00	3.50	34.50	21.50	24.50	8.50	2.00

Table 4.2: Percentages of tag types used for web resources in *Delicious*, images in *Flickr*, and music in *Last.fm*

4.1.3 Tagging Systems and Tag Type Prevalence

In order to determine the prevalence of each tag type in the different collaborative tagging systems, we manually classified 2,100 tags, 700 for each system. These tag samples were drawn in the following manner: we took the 300 top tags, then 200 tags from where 70% of the probability density starts, and again 200 tags starting from 90%. The goal is to capture the distinct patterns the various parts under the power law distribution, i.e., popular, less popular, and rather idiosyncratic tags, exhibit (see also [101]). Since our three datasets are not comparable in terms of absolute size, the cut-off happens at different points. Nevertheless, the resulting effect should be insignificant as the tail mostly contains tags used very infrequently. Please note that in rare cases tags had to be skipped (and replaced) as they were unreadable or their meaning could not be understood.

Table 4.2 shows the results of our intellectual classification[4]. The bold figures correspond to the percentages found for the total sample of 700 tags per system. The other three figures per system report the numbers for the subsamples of varying tag popularity[5]. Most noticeable is that distribution of tag types varies across collaborative tagging systems with high significance ($p < 0.001$) according to a Chi-Square

[4]The distributions found strongly resemble those reported in [23] for a smaller subsample of 300 tags per system. One notable difference can be observed though: For *Last.fm*, the percentage of Type tags decreases given the bigger sample covering also less popular tags. This may be explained by the prevalence of – an restricted set of – genre tags among the top popular tags.

[5]Due to the difference in subsample size, the percentage for the total sample in each system is not (exactly) the average over its three subsamples.

test of homogeneity[6] (see, e.g. [79]). This statistical test belongs to the family of Chi-Square tests introduced by Karl Pearson. It is used for determining whether frequency counts for a categorical (also nominal) variable have the same distributions in two different populations [79].

While for *Delicious* and *Flickr* Topic is the most prominent category with more than 50% of the tags, it is Type for *Last.fm* – mainly due to genre tags. Thus, music is more typically described and organized via genre labels and only rarely by topic (e.g. "love"). For websites and pictures, in contrast, organization around topics is natural since there exist so many web page topics and picture motives. Type is the second most frequent tag category used in *Delicious*, indicating that web resources cover a great variety of media formats as well as text genres. Since for photos in *Flickr* Type tags name rather professional photographic techniques and styles (e.g., lens type, "macro"), they are not used heavily by users. What is important for characterizing pictures, is Location. Especially among the top tags in *Flickr* many tags name a place associated to the photo. Similarly, [105] report that photos in *Flickr* are preferably tagged with content, location, and device.

Usage context is used to a comparable amount of 8% in *Delicious* and *Flickr* and with 6% in *Last.fm*. It is striking that for *Last.fm* subjective opinions are way more important. This may be attributed both to the free-for-all-tagging character – with less motivation for personal information management – and its focus on music – with music playing a major role in (displaying) young people's self-concepts and developing identity (see also [239, 181]). Time and Self-reference are rarely used in either system. Author/Owner tags appear a little more often in *Last.fm*, naming track artists, and in *Delicious*. Since tagging in *Flickr* is more or less concerned with tagging own pictures [155], this category is practically absent here.

Regarding variation of tag type prevalence with respect to tag popularity, Table 4.2 shows that the importance of the different tag types remains more or less stable over the subsamples for each system. Topic stays the dominant tag type for *Delicious* and *Flickr* and Type for *Last.fm*. For *Delicious*, Author/Owner tags increase at the expense of Type tags. However, the differences found for the second and third subsample of less popular tags are not significant (Chi-Square test of homogeneity; rejected since $p > 0.05$). From the top tags both samples differ significantly ($p < 0.01$). For *Flickr*, saliency of Topic tags increases steadily with decreasing tag popularity. Time and Topic, on the other hand, loose prominence. Here, the

[6] Categories having a frequency less than five in either system of the (pairwise) comparison were ignored in the test, i.e., they were removed from the contingency table as it has been argued that the Chi-Square statistic is not reliable for small values [140].

more limited vocabulary for the usually rather coarse-grained time expressions (year, month, weekday) may be a partial explanation. Again, the frequency distributions in the three subsamples differ significantly from each other ($p < 0.05$ and lower).

In *Last.fm*, Location and Type tags are less frequent among less popular tags while Author/Owner and Opinion/Qualities tags become considerably more prevalent. As the latter often consist of longer words or even phrases (self-expression), it is not surprising that convergence or agreement on such tags is lower than, e.g., on genre tags. These differences in the distribution of tag types in the long tail sample are highly significant ($p < 0.001$). The top tags and the second sample of tags with medium popularity, though, do not show significantly different frequencies ($p > 0.05$).

Summing up, more than 80% of the tags in *Flickr* and *Delicious* are factual in nature and thus verifiable. For *Last.fm*, the corresponding figure is with 71% still very high. Thus, such annotations describing a resource by considerung different non-subjective aspects are usable by other users as well. They are not only beneficial to the tagger herself/himself. The two subjective categories Opinion/Qualities and Usage context together make up for only 13% of all sample tags from *Delicious* (15% for *Flickr*), and they may still be helpful for our purposes of building tag profiles to match potentially like-minded people or to enable better search and social routing through tags. For *Last.fm*, with its large amount of subjective Opinion/Qualities tags, a somewhat more elaborated strategy to deal with these tags may be appropriate (e.g., after automatic identification). Such tags may be highly ideosyncratic and even top tags in this category may not reflect well less popular opinions. Less than 5% are truly personal (Self-reference) tags, hardly useful for user profiling or enhanced search.

Besides, our analysis could establish an additional value of tags (for details please refer to [23]): More than half of the tags do not appear in the content of the web resources they annotate and, thus, carry new information (see also [106] for similar results). For multimedia data like music or pictures, tags have an even higher benefit, since the data is not easily interpreted semantically (so called semantic gap). For example, only 1.5% of the tags were found when looking up tags of music tracks in the lyrics of the song. Hence, 98.5% of the tags add exploitable knowledge in form of text strings, which can be used to enhance textual search or mapping of user profiles. Moreover, a large amount of tags is also representative, accurate, and reliable. In the music domain, for example, 46.14% of the tags also occur in online music reviews written by experts on *AllMusic*. Even 73.01% appear in web reviews retrieved from Google by posting a query of the form ["artist" "track" music review -lyrics].

	Topic	Time	Location	Type	Author/ Owner	Opinions/ Qualities	Usage context	Self-reference
Web	**35.67**	**0.83**	**6.78**	**3.26**	**47.11**	**4.79**	**1.23**	**0.33**
top 100	18.00	1.00	0.00	6.00	73.00	1.00	0.00	1.00
2nd 100	29.32	0.75	12.03	2.26	54.89	0.75	0.00	0.00
3rd 100	60.00	0.74	8.89	1.48	12.59	12.59	3.70	0.00
Pictures	**50.13**	**6.18**	**4.67**	**4.78**	**4.56**	**24.34**	**5.05**	**0.28**
top 100	49.15	0.85	0.85	5.08	6.78	32.20	4.24	0.85
2nd 100	47.24	17.79	2.45	7.98	4.91	14.72	4.91	0.00
3rd 100	53.06	0.00	10.88	1.36	2.04	26.53	6.12	0.00
Music	**15.97**	**3.39**	**3.31**	**15.42**	**17.52**	**15.65**	**28.22**	**0.52**
top 100	3.42	1.71	3.42	28.21	13.68	18.80	29.91	0.85
2nd 100	27.01	1.46	5.11	13.87	27.01	4.38	21.17	0.00
3rd 100	17.93	6.90	1.38	4.14	12.41	23.45	33.10	0.69

Table 4.3: Frequencies of query types found in the AOL log regarding search for web resources, images, and music

4.1.4 Correspondence between Tagging and Search Behavior

To be truly useful for social search, search queries and tags need to match. Therefore, we report an analysis on the AOL [176] web query log. In a first experiment we found that 71.22% of general web queries contain minimum one tag used in *Delicious*. For 30.61% of the queries, all query terms can be matched to a *Delicious* tag. The corresponding figures are 64.54% and 12.66% for *Flickr* and 58.43% and 6% for *Last.fm*. We have to keep in mind, though, that our *Delicious* sample vocabulary is much bigger due to the crawling methods employed for data collection. Nevertheless, this considerable overlap makes the benefit of user generated tags obvious (see similar results for *Delicious* in [106]).

Regarding types of query terms, we also classified sample queries from the log file into our eight tag categories and counted their frequencies. For differentiating music and picture queries from general web queries, queries were selected that contained words like "song", "music", "photo", or that led to a click on *Flickr* or *Last.fm*. Lists sorted by query frequency were created. Similar to the procedure for tags, three subsamples of varying popularity were then taken from each list. A query could be assigned to multiple categories, once for each distinct category one or more of its terms fell into. Table 4.3 shows the resulting distributions, again with figures in bold for the entire sample per resource type. The distribution of query types also varies significantly across resource types ($p < 0.001$; Chi-Square test of homogeneity).

As expected, most web queries ask for the topic of web pages to be searched for. The frequencies per category are comparable to the ones for tagging in *Delicious* except for the somewhat less frequent Topic queries and Author/Owner tags/queries. While the latter are rarely found in *Delicious*, quite a few web queries fall into that category. Most of these queries are navigational queries, asking for highly popular sites like search engines (e.g., "google", "yahoo.com"), online shopping (e.g., "ebay", "amazon"), email providers (e.g. "hotmail"), etc., as an alternative to direct access, i.e., typing the URL into the browser[7]. Frequencies of these two categories are significantly different in *Delicious* and our web queries sample ($p < 0.001$). With decreasing query popularity informational Topic queries become more prominent at the expense of such queries for Author/Owner. Distribution differences between the subsamples are all significant ($p < 0.05$ or lower).

With respect to queries for pictures, Topic is with around 50% the most important category – so it is in *Flickr*. The relative category frequencies for queries strongly resemble the frequencies of tags in *Flickr*. However, there are two notable differences regarding Location, being searched less, and Opinions/Qualities. Such subjective queries ("funny", "sexy") are quite prominent in our picture samples mainly due to the abundance of 'adult' queries. Therefore, the distributions of tags types in *Flickr* and the distribution of picture queries differ significantly ($p < 0.001$). However, category proportions did not depend on query popularity ($p > 0.05$).

Searching and tagging music shows some distinct characteristics ($p < 0.001$). There is also a highly significant dependence of query type frequency on popularity ($p = 0$). While Type – mainly covering genre labels – is the dominant tag type in *Last.fm*, it is considerably less often used to search for music. For less popular queries, Type (genre) queries are hardly used at all. Two explanations seem plausible for the reduced value of genres in public search for music. First, genre may not be distinctive enough, thus, potentially retrieving too many hits. On the other hand, genre classification is to a certain degree ambiguous in that it depends on the musical expertise of the person annotating as well as on the (imperfect and continuously evolving) genre classification scheme used.

Instead, users search for known music by providing the title and artist/author of a song to be retrieved. Tags of this kind are rare probably due to *Last.fm* providing this information in the metadata for each song. Thus, as support for future retrieval by Title and Artist is given by default in the platform, tagging this information

[7] Broder [36] had identified different information needs behind users' web queries and proposed a taxonomy consisting of navigational (go to a particular site), informational (gather information), and transactional queries (perform a web-based activity like online banking).

does not offer added value. Users also often search for music that fits a specified usage context (e.g. "wedding music") and, in particular, for music appearing on the soundtrack of a movie, a video game, etc. Almost identical is the proportion of subjective Opinion/Qualities tags and queries.

30 PhD students in computer science (23 male, seven female), aged between 23 and 40, participated in a study where they rated the usefulness of each of our tag categories for a.) searching personal resources and b.) searching public resources of other users. Participants were also asked to judge how well each category can be remembered. All ratings were given on a five-point Likert scale with responses in the range from zero (not useful/bad to remember) to four (very useful/very good to remember).

Ratings do highly correlate for the three different activities of personal collection search, public search and remembering (pairwise correlation coefficients are between 0.85 up to even 0.97), with minor deviations for Time and Type for pictures and Opinions/Qualities for music. Time, for example, is more valued and well remembered for one's own resources. For resources on the web, time may not be known at all. For images, on the other hand, the Type category is helpful for narrowing down public search while it is not that useful in one's personal collection – probably since personal pictures are so rarely annotated with such professional tags.

However, ratings on categories do vary across resources types. For web pages, Topic followed by Usage context, Author/Owner, and Type is rated best regarding usefulness and remembering. For pictures, Topic, Location, and Time are rated most valuable – similar to the importance of those categories for tagging in *Flickr*. For music, Author/Owner, Type, and Topic were rated high. Opinions/Qualities are judged more useful for searching songs on the web, less for searching favorite songs in your personal collection.

Not only for music, it seems that people assume quite some agreement on subjective characteristics and opinions. Though users considered the 'factual' categories more useful, they also valued the more subjective categories Opinions/Qualities and Usage context surprisingly high. We already reported on findings [165] stating that people turn to their social networks for subjective queries to get recommendations. In the case of tags, "trust" in the appropriateness of subjective tags seems still to be present despite a missing social connection.

We will now give further evidence on the potential and the quality of user generated tags, namely by deriving new knowledge about resources based on given tags.

4.2 Knowledge Mining from Tags: Mood and Theme Identification for Music

Our analysis showed that big part of music queries falls into the Usage context and Opinion/Qualities categories: Almost 30% of the queries are theme-related and 16% target mood information. However, such tags – themes in particular – are underrepresented among popular tags in the tagging system *Last.fm*. Hence, our goal is to automatically derive mood and theme metadata for music tracks to better cover diverse facets reflecting the complex real-world music information needs of users. With the "mood of a song" we denote the state or the quality of a particular feeling induced by listening to that song (e.g., "aggressive", "happy", "sad"). The "theme of a song" refers to the context or situation which fits best when listening to the song, e.g., "at the beach", "night driving", "party time".

Currently available state-of-the-art music search engines still do not explicitly support music retrieval based on mood and theme information, and content-based approaches trying to address this problem mainly focus on identifying the moods of songs and do not tackle the thematic aspects of the music resources. Several works in Music Information Retrieval (MIR) have shown a potential to model the mood from audio content (like [147, 72, 202, 233], see [131] for an extensive review). Although this task is quite complex, satisfying results can be achieved if the problem is reduced to simple models [131]. However, an important limitation of these approaches is that they concentrate on the mood only expressed in the audio signal itself. Thus, they can not capture other sources of emotionality.

Apart from analyzing the low-level features of music resources to identify the songs' corresponding mood or theme, we will show that collaboratively generated tags are a powerful source of information that can be used for the task. We use *Last.fm*'s valuable folksonomy information for inferring mood and theme labels for songs. While in earlier experiments only tags were used for deriving moods, themes, and styles/genres [24, 26], here we also investigate fusion with audio-based methods[8]. Extending existing music metadata enrichment studies, we fuse social tags and low-level audio features of the tracks to infer mood or theme labels, showing that both sources provide helpful complementary information.

The contributions of the following experiments are twofold:

- We show the feasibility of automatic music classification according to contextual aspects like themes.

[8]The experiments are joint work with colleagues from the Music Technology Group at the Universitat Pompeu Fabra, who provided all audio data as well as the audio-based classification algorithms.

- We successfully exploit collective knowledge in form of tags in order to complement the intrinsic information derived from audio features.

The algorithms can be used, for example, to index predicted mood and theme labels to enrich the metadata index of music search engines enabling a more social and context-aware search (or browsing). Besides, such labels will be valuable for recommendation and playlist generation, e.g., for listening to "Party Time" music.

4.2.1 Datasets

AllMusic. *AllMusic* (AMG), established in 1995, is a community website for music fans. There one can find information on music tracks, albums, and artist covering plenty of music genres and styles. Amongst others, there are also reviews of albums and artists as well as labels for instruments, moods, and themes related to a song. Since both reviews and classifications are manually provided by music experts from the *AllMusic* team, the data can be considered a valid ground truth. For the experiments, we gathered 178 different moods and 73 themes from the *AllMusic* pages corresponding to music themes and moods. From these pages we also collected information on which music tracks are classified according to these categories. Our final song set comprises 5,770 songs. For these songs, we have 8,158 track-mood and 1,218 track-theme assignments. On average, tracks are annotated with 1.73 moods and 1.21 themes, with a maximum of 12 and six annotations respectively.

Last.fm. For the tracks collected from *AllMusic*, we obtained the *Last.fm* tags users had assigned to these songs together with the corresponding frequencies. As not all *AllMusic* songs have user tags in *Last.fm*, our set of tracks is reduced to 4,737. Using the API[9], we collected in total 59,525 different tags for this set of songs.

Audio. For each track from the previous two collections, we did a lookup in the audio database of our colleagues from the Music Technology Group at the Universitat Pompeu Fabra. For those songs present, our colleagues automatically extracted several state-of-the-art MIR audio features from the 30 seconds excerpts (mp3 format with a bit rate of 192 kbps): timbral, tonal, rhythmic including MFCCs, BPM, chroma features, spectral centroid, and others. Please refer to [131] for a complete list. For each excerpt of the data set, its 200ms frame-based extracted features were summarized with their component-wise means and variances. At the end of the process, we obtained 240 low-level and mid-level audio features.

[9] `http://www.last.fm/api`

4.2.2 *AllMusic* Class Clustering

Given that the number of classes existing in *AllMusic* is quite large (e.g., 178 different moods), with many of the individual terms being highly synonymous or denoting the same concept in well known models of emotions[10] [112], clustering was applied to the initial sets of *AllMusic* moods and themes.

4.2.2.1 Mood Clustering

For comparison reasons, we choose the five mood categories used for the MIREX Audio Music Mood Classification Track (see Table 4.4)[11]. Each of the clusters consists of five to seven *AllMusic* mood labels that together define the cluster. These categories were proposed in [112], derived from a popular set (of top songs, top albums). The MIREX mood clusters seek to reduce the diverse mood space while still capturing the social-cultural context of pop music. Restricting our data set to tracks whose assigned moods fall into exactly one of these categories, we had 1,192 distinct songs left for machine learning. To balance cluster size for our multiclass classifiers, the cut-off was set to 200 instances per cluster. The original cluster sizes were 214 for MM1, 205 for MM2, 300 for MM3, 273 for MM4, and 200 for MM5.

Since many *AllMusic* mood labels and thus the corresponding songs classified by human experts are not used in MIREX, we as well experimented with the well known two-dimensional models of emotion/mood. In the Thayer energy-stress model [209], emotions are classified along the two axes of (low - high) energy and (low - high) stress. Thus, the two factors divide the mood space into the four clusters "exuberance", "anxious/frantic", "depression", and "contentment". Similarly, Russell/Thayer's bipolar model differentiates emotions based on arousal and valence. In the psychological literature there is little agreement on the number of basic emotional categories or dimensions – it is unclear how many different moods people really distinguish in terms of linguistic description, physiological reaction, etc. However, the Thayer model has been proven useful for music classification, and the four categories resulting seem a fair compromise: reducing the mood space to enable clear classificatory distinction and still providing valuable extra-musical metadata for exploratory information needs.

During clustering we manually mapped all *AllMusic* labels into the two-dimensional mood space by adopting a similarity sorting method as described below for themes.

[10] Moods are highly related to emotions though they differ in duration, are less intensive, and lack object directedness.

[11] `http://www.music-ir.org/mirex/wiki/2007:Audio_Music_Mood_Classification`

No.	MOOD CLUSTERS – MIREX
MM1	Passionate, Rousing, Confident, Boisterous, Rowdy
MM2	Rollicking, Cheerful, Fun, Sweet, Amiable/Good natured
MM3	Literate, Poignant, Wistful, Bittersweet, Autumnal, Brooding
MM4	Humorous, Silly, Campy, Quirky, Whimsical, Witty, Wry
MM5	Aggressive, Fiery, Tense/Anxious, Intense, Volatile, Visceral
No.	**MOOD CLUSTERS – THAYER**
MT1	**high energy / high stress:** Druggy, Raucous, Paranoid, Manic, Brittle, Fiery, Spooky, Eerie, Rowdy, Angry, Fierce, Aggressive, Rebellious, Trippy, Brash, Provocative, Boisterous, Thuggish, Hostile, Angst-Ridden, Volatile, Enigmatic, Harsh, Ominous, Rambunctious, Malevolent, Menacing, Reckless, Unsettling, Confrontational, Theatrical, Outrageous, Uncompromising, Tense/Anxious
MT2	**high energy / low stress:** Rollicking, Exuberant, Crunchy, Sexy, Exciting, Searching, Sparkling, Summery, Party/Celebratory, Witty, Intense, Visceral, Energetic, Spicy, Ambitious, Giddy, Sensual, Happy, Gleeful, Sexual, Gutsy, Spacey, Humorous, Epic, Lively, Swaggering, Organic, Cheerful, Hedonistic, Fun, Rousing, Bravado, Freewheeling, Carefree, Passionate, Earthy, Playful, Gritty, Joyous, Amiable/Good-Natured
MT3	**low energy / low stress:** Calm/Peaceful, Stylish, Lush, Sophisticated, Soft, Sentimental, Refined/Mannered, Cathartic, Romantic, Springlike, Smooth, Warm, Precious, Ethereal, Confident, Hypnotic, Naive, Intimate, Cerebral, Indulgent, Innocent, Reverent, Literate, Relaxed, Soothing, Slick, Earnest, Dreamy, Gentle, Sweet, Elegant, Laid-Back/Mellow, Light
MT4	**low energy / high stress:** Melancholy, Quirky, Detached, Delicate, Irreverent, Restrained, Brooding, Whimsical, Campy, Sparse, Meandering, Sad, Gloomy, Snide, Somber, Autumnal, Weary, Wry, Wintry, Plaintive, Nocturnal, Clinical, Poignant, Yearning, Wistful, Austere, Bittersweet, Reserved, Cynical/Sarcastic, Fractured, Bleak, Reflective, Ironic, Bitter, Acerbic, Silly, Sardonic

Table 4.4: Mood clusters

Table 4.4 shows the four resulting clusters together with example *AllMusic* labels. Again, clusters were balanced by randomly choosing 403 instances for each cluster – the size of the smallest cluster.

4.2.2.2 Theme Clustering

Since *AllMusic* themes do not directly correspond to human emotions, mapping the 73 theme terms into the mood spaces used before was not possible – though themes may often be associated with specific moods. For manual clustering, we adopted a similarity sorting procedure like in [197], which we also used in prior related work [24]. For this, all *AllMusic* themes were written on pieces of paper, which were then grouped by building piles of themes judged as belonging together. Final groupings were derived by analyzing the clusters showing up in co-occurrence matrices. Extensive discussions helped resolving cases of ambiguity regarding term membership. This procedure yielded a theme list comprising 13 labels. Classes containing too few songs are discarded in order to have a minimal representative learning corpus for the classifier. As a result, the remaining four theme clusters (Table 4.5) contain 74 songs each.

No.	**THEME CLUSTERS**
T1	Party Time, Birthday Party, Celebration, Prom, Late Night, Club, Guys Night Out, Girls Night Out, At the Beach, Drinking, Cool & Cocky, Pool Party, Summertime, TGIF (Thanks God It's Friday)
T2	Sexy, Seduction, Slow Dance, Romantic Evening, In Love, New Love, Wedding, Dinner Ambiance
T3	Background Music, Exercise/Workout Music, Playful, Day Driving, Victory, The Sporting Life, Long Walk, The Great Outdoors, Picnic, Road Trip, Motivation, Empowering, Affirmation, The Creative Side, At the Office
T4	Divorce, Heartache, Feeling Blue, Breakup, Regret, Loss/Grief, Jealousy, Autumn, Rainy Day, Stay in Bed, Sunday Afternoon, Solitude, Reminiscing, Introspection, Reflection, Winter

Table 4.5: Theme clusters

4.2.3 Classification Algorithms

For predicting themes and moods, we base our solution on social knowledge – i.e., collaboratively created tags associated to music tracks – extracted from *Last.fm* as well as on audio information. Building upon already provided user tags, on the audio content of music tracks, or on combinations of both, we build multiclass classifiers to infer additional annotations corresponding to moods and themes.

The core of our mood and theme classification methods are multiclass classifiers trained on the *AllMusic* ground truth using tags or audio information as features. We experiment both with classifiers created separately for the two different types of features we consider, which are then combined in order to produce for each song a final mood/ theme classification, as well as with a classifier taking as input a combination of audio and tag features. After several experiments, we could observe that the libsvm[12] implementation [46] of Support Vector Machine (SVM) classifiers with Radial Basis Function (RBF) kernel performed best for the case of audio input

[12] A library for Support Vector Machines: `http://www.csie.ntu.edu.tw/~cjlin/libsvm`

features. The basic idea behind the very popular SVM classifiers (see, e.g. [59]) is to find the maximum-margin hyperplane, which best separates the data points from two given classes. Besides directly operating on the dimensions spanned by the features of a resource (linear SVM), data points can also be mapped via non-linear kernel methods into a transformed feature space of higher dimensionality to better enable linear separation. SVM classifiers can also be applied to multiclass problems by training multiple two-class (one-against-one/one-versus-one) classifiers or one-against-all (one-versus-all) SVMs.

In the case of tag features, WEKA's implementation of Naïve Bayes Multinomial [159] achieved the best performance. In short, Naïve Bayes classifiers estimate the probability of a document belonging to a certain class based on the probabilities the document's single features have for co-occurring with that class as well as the prior probability (or base rate) of the class. The approach has proven well suited for text classification tasks despite the 'naïve' assumption of feature independence.

Additionally, the linear combination of the separate classifiers for audio and tag features performed better than the classifier trained on audio and tag features simultaneously. Only the best obtained classification results are presented here. We have classifiers trained for the whole set of classes (i.e., either for moods or themes), and these classifiers produce for every song in the test set a probability distribution over all classes (e.g., over all moods). The highest probability is considered in order to assign the songs to the corresponding class. We experimented with feature selection based on automatic methods like Information Gain, but the results showed that the full set is better suitable for learning even though it contains some noise.

Algorithm 4.1 presents the main steps of our classification approach, where classifiers are trained separately for the two different types of input features – tags and audio information. We show the algorithm for mood classification, the case of theme classification being similar.

Step 1 (optional) of the algorithm above aims at reducing the number of mood classes to be predicted for the songs. If two classes are clustered, the resulted class will contain all songs which have been originally assigned to any of the composing classes. As we need a certain amount of input data in order to be able to consistently train the classifiers, we discard those classes containing less than a certain number of songs[13] assigned (Step 2). After selecting separate sets of songs for training and testing in Step 3a, we build the feature vectors corresponding to each song in the training set (Step 3b). In the case of features based on tags, each distinct tag

[13]The exact numbers depend on clustering / class type and are given in Section 4.2.2.

Alg. 4.1. Mood classification

Input:ftype – feature type

$$ftype = \begin{cases} 0, \text{for tag features;} \\ 1, \text{for audio features.} \end{cases}$$

M – mood classes to be learned

S_{total} – set of songs

1: *Apply clustering method to cluster moods (see Section 4.2.2)*

2: *Select classes of moods M to be learned*

For each mood class

  If the class does not contain at least X songs

    Discard class

3: *Classifier learns a model*

3a: Split song set S_{total} into

  S_{train} = songs used for training the classifier

  S_{test} = songs used for testing the classifiers' learned model

3b: Select features for training the classifier

If ($ftype = 0$) // tag features

  For each song $s_i \in S_{train}$

    Create feature vector $F_t(s_i) = \{t_j | t_j \in T\}$, where

    T = set of tags from all songs in all mood classes

$$t_j = \begin{cases} log(freq(t_j) + 1), \text{if } s_i \text{ has tag } t_j; \\ 0, \text{otherwise.} \end{cases}$$

Else // audio features

  For each song $s_i \in S_{train}$

    Create feature vector $F_a(s_i) = \{a_j | a_j \in A\}$, where

    A = set of audio features from all songs in all mood classes

    $a_j = standardize(a_j)$

3c: Train and test classifier

If ($ftype = 0$) // tag features

  Train Naïve Bayes (NB) on S_{train} using $\{F_t(s_i); s_i \in S_{train}\}$

  Test Naïve Bayes (NB) on S_{test}

Else // audio features

  Train SVM on S_{train} using $\{F_a(s_i); s_i \in S_{train}\}$

  Test SVM on S_{test}

4: *Classify songs into mood classes*

For each song $s_i \in S_{total}$

If ($ftype = 0$) // tag features

  Compute probability distribution $P_t(s_i)$ as

  $P_t(s_i) = \{p_{NB}(m_j|s_i); m_j \in M\}$

  Assign s_i to m_j, where $max(p_{NB}(m_j|s_i))$

Else // audio features

  Compute probability distribution $P_a(s_i)$ as

  $P_a(s_i) = \{p_{SVM}(m_j|s_i); m_j \in M\}$

  Assign s_i to m_j, where $max(p_{SVM}(m_j|s_i))$

---

assigned to the songs belonging to the mood classes makes up one element in the feature vector. The elements of a vector will have values depending on the frequency of the tags occurring along with the song. We experimented with different variations for computing the vector elements, but the formula based on the logarithm of the tag frequency provided best results. Audio features are standardized for better suitability with the SVM classifier. Here, a one-vs-one multiclass approach was taken with the parameters selected via grid search (C and gamma with three-fold cross validation method). Probability estimations are made by pairwise coupling [227]. Once the feature vectors are constructed, they are fed into the classifier and used for training. After the model is learned, it is applied in order to produce predictions of the songs belonging to the different mood classes. The assignment of a song to a class is done based on the maximum predicted probability among all possible classes (Step 4).

As already mentioned, we also experiment with a linear combination of the predictions of the two separately trained classifiers (details are presented in Algorithm 4.2). The two different classifiers are first trained to make predictions for all songs in the collection (Step 1). For producing a linear combination of the classifiers as final output, we then experiment with different values of the α parameter. We choose the α value for which the maximum $F1$ is achieved (see evaluation section below). We then use it within our linear combination of the audio and tag-based classifiers in order to produce the assignment of songs to the mood classes.

Alg. 4.2. Mood classification – classifiers' linear combination
Input: M – *mood classes to be learned*
S_{total} – *set of songs*

1: For each song $s_i \in S_{total}$
Compute $P_a(s_i) = \{p_{SVM}(m_j|s_i)\} = \{p_a(m_j|s_i)\}$
and $P_t(s_i) = \{p_{NB}(m_j|s_i)\} = \{p_t(m_j|s_i)\}$ (see Alg. 1, step 4)
2: For each α=0.1,...,0.9, *step*=0.1
For each song $s_i \in S_{total}$
For each mood $m_j \in M$
$p_{at}(m_j|s_i) = \alpha \cdot p_a(m_j|s_i) + (1-\alpha) \cdot p_t(m_j|s_i)$
Assign s_i to m_j, where $max(p_{at}(m_j|s_i))$
Compute P, R, Acc, $F1$
3: Select $\alpha = \alpha_{best}$ that produces best results for P, R, Acc, $F1$
4: Classify songs into mood classes, using α_{best} for weighting the probabilities output by the audio-based classifier and $(1 - \alpha_{best})$ for weighting the probabilities predicted by the tag-based classifier.

Classifier	**Class**	**R**	**P**	**F1**	**Acc**
SVM (audio)	Mood MIREX	0.450	0.442	0.420	0.450
NB (tags)	Mood MIREX	0.565	0.566	0.564	0.565
Comb. ($\alpha = 0.7$)	Mood MIREX	0.575	0.573	0.572	0.575
SVM (audio)	Mood THAYER	0.517	0.515	0.515	0.517
NB (tags)	Mood THAYER	0.539	0.542	0.539	0.539
Comb. ($\alpha = 0.8$)	Mood THAYER	0.570	0.569	0.569	0.569
SVM (audio)	Themes clustered	0.528	0.581	0.522	0.527
NB (tags)	Themes clustered	0.595	0.582	0.575	0.595
Comb. ($\alpha = 0.9$)	Themes clustered	0.625	0.617	0.614	0.625

Table 4.6: Recall R, Precision P, F1-Measure $F1$, and Accuracy Acc for the different classifiers, moods, and themes

4.2.4 Evaluation

For measuring the quality of our theme and mood predictions, we compare our output against the *AllMusic* experts' assignments, using Precision (P), Recall (R), Accuracy (Acc), and F1-Measure ($F1$) for the evaluation [183, 152]. Precision measures the ratio of relevant documents retrieved while Recall gives the ratio of relevant documents actually retrieved over all relevant documents. The F1-Measure combines both metrics to derive a harmonic mean capturing specificity as well as completeness simultaneously. Accuracy is an alternative metric counting the percentage of correct classifications, i.e., of all true positives and true negatives.

We present the best results achieved among all our experimental runs (10-fold cross validations) in Table 4.6. These runs correspond to the different combinations of classifiers (audio-based, tag-based, or linear combinations of the two) and classes to be predicted (themes or moods clustered according to MIREX or Russell/Thayer).

For both moods and themes, we observe that the classifiers relying solely on audio features perform worse than the pure tag-based classifiers. However, combining the two types of classifiers leads to improved overall results. For the moods clustered according to MIREX, Russell/Thayer, and for themes manually clustered, the best values of α are 0.7, 0.8, and 0.9 respectively. These values indicate a higher weight for the audio-based classifiers though their achieved performance is poorer than that of the tag-based classifiers. This fact is easily explainable by the different types of classifiers considered: SVM for audio features and Naïve Bayes for tag features. It is known that Naïve Bayes produces probabilities close to one for the most likely identified class whereas for the rest of classes the probabilities are closer to zero.

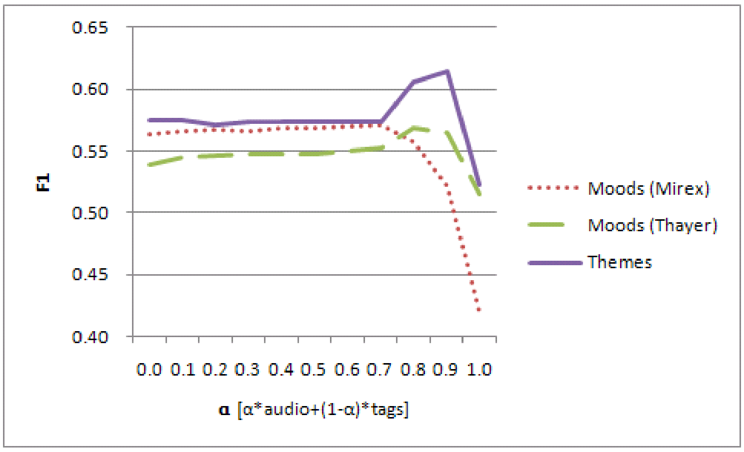

Figure 4.1: $F1$ values for varying α when linearly combining audio-based and tag-based mood and theme classification

SVM, in contrast, produces more even probability distributions. Therefore the high probabilities output by Naïve Bayes need to be evened out through a lower α weight.

The variations of the $F1$ measure with α are depicted in Figure 4.1. The biggest variations are to be found in the case of moods clustered into MIREX mood classes, where for *alpha* values of 0.7 and higher we observe a sharp drop of the $F1$ value. For the Russell/Thayer model, the $F1$ values start to deprecate with α values greater than 0.8.

The baseline accuracy for a random classifier trying to assign songs to the Russell/Thayer mood classes or to the theme clusters is 0.25 while for the MIREX mood classes it would be 0.2. The linear combination of the classifiers improves accuracy in the range of 10 to 27.7% for moods and 18.5% for themes over audio-based classifiers. Overall, results are better for theme classification, indicating that themes are easier to distinguish. Analyzing the confusion matrices for the best performing approaches (Figure 4.2), we observe some prominent confusion patterns: In the case of MIREX clustering, instances belonging to class $MM1$ are often misclassified into $MM2$, $MM4$ instances into $MM3$. Similarly, $MT3$ instances are wrongly classified into $MT4$ for the case of Russell/Thayer clustered moods; also $MT1$ and $MT4$ are often confused. For the latter, the energy dimension does not seem to ease differentiation, given that high stress (negative valence) is characteristic for both classes. $T3$ and $T1$ are the difficult theme classes. Further refinement of these classes should be considered for future work in order to eliminate this kind of ambiguities (e.g., “Exercise/Workout” music might be as well considered “Party Time” music).

It is difficult to directly compare our results to the related work cited as each paper uses a different number of classes. Moreover, experimental goals, ground

A) Moods (Mirex)

Correct Class \ Predicted Class	MM1	MM2	MM3	MM4	MM5
MM1	83	49	14	24	30
MM2	26	118	33	21	2
MM3	6	24	134	27	9
MM4	14	23	38	101	24
MM5	25	6	10	20	139

B) Moods (Thayer)

Correct Class \ Predicted Class	MT1	MT2	MT3	MT4
MT1	269	50	30	54
MT2	57	229	48	69
MT3	40	80	198	85
MT4	61	38	82	222

C) Themes

Correct Class \ Predicted Class	T1	T2	T3	T4
T1	39	8	13	14
T2	3	60	6	5
T3	19	12	30	13
T4	4	9	5	56

Figure 4.2: Confusion matrix for the best approaches

truth, and evaluation procedures vary as well, or detailed descriptions are missing. Comparing to the best algorithms submitted to the MIREX task, we achieve results with lower accuracy. However, knowing that the algorithm used in this paper for audio classification is the same as submitted to MIREX in 2007 [130] (obtaining 60.5% accuracy), our conclusion is that the difference comes from the ground truth data. The hypothesis is that our results here are lower because we did not filter the training and test instances using listeners. Moreover, for the MIREX collection, listeners were asked to focus on audio only (not lyrics, context, or other), which makes it much easier then to classify using audio-based classifiers. In that context, the classification task on our MIREX-like *AllMusic* ground truth is more difficult.

In [24], we contrasted tags with lyrics as another source of information possibly useful for categorizing music according to *AllMusic* genre, styles, and again moods, and themes. For the latter, we also experimented with different clustering techniques (manual, co-occurrence-based, via WordNet). With respect to moods, we made use of a different classification of human emotions: the hierarchy presented in Shaver et al. [197]. This model comprises six primary emotions ("Fear", "Sadness", "Joy", "Love", "Anger", "Surprise"), each of which is the superclass of one to six more complex secondary emotions.

While for genre and style recommendation tags alone led to best performances of hit rate @3 of 0.97 and 0.76 respectively, for themes and moods success of using only tags and combining tags with lyrics was comparable (hit rate @3 of 0.88 for themes clustered via WordNet, 0.89 for primary emotions, and 0.64 for secondary emotions. Hit rate at rank k is defined as the probability of finding a good descriptive tag among the top-k recommended tags. We also conducted a user evaluation in form of a *Facebook* application, where users could select one or more theme labels they thought fit a particular song. Taken together, these results indicate the usefulness of tags for mining additional knowledge – leading to superior performance when taken as single source of input as well as improving performance of content-based approaches by combining these with the 'collective intelligence' captured in tags.

4.3 Discussion

User generated tags hold various potentials for improving user profiling and search. First, the results of our analysis on tagging behavior in the different collaborative tagging systems *Delicious*, *Flickr*, and *Last.fm* showed that though the dominant tag type (e.g., topic vs. genre) varies across systems the majority of tags comes from factual, thus verifiable, tag categories. These tags seem also pretty reliable as they resemble expert labels. In addition, many tags hold new information about a resource, and do not simple copy what is already in the page text. For multimedia, the gain is substantial as usually no or only few textual information is available that capturing content characteristics or subjective opinions.

Interestingly, from the frequency of subjective search engine queries found and from the results of our small user study we find that subjectivity is still valued especially for unknown multimedia resources and that obviously agreement is assumed to a certain degree. Such correspondence in tagging behavior is crucial for using tags in user profiling and matching, for example, for supporting tie formation with like-minded people or for enhancing search by extending it over the network of such similar users. Similarly, search and recommendation on tags in general needs quality tags capturing semantics of resources through collective intelligence. Since users' tagging behavior is also to a big part in accordance with their querying patterns, tags can be considered beneficial for improving search. Less than 5% of the tags were self-references and, thus, hardly benefitial for neither user profiling and matching nor search.

Regarding user profiling, tags are also a valuable source for matching user identities across different collaborative tagging sites. In one of our studies [115], given alone the tag profile from one tagging system (e.g. *Delicious*) the user can be identified within the other one (e.g. *Flickr*) with a hit rate @1 of around 30-34%. Combining tagging information with username matching via a mixture model the success rate (@1) improves to even 64%.

The potential of social tags is proven once more through experiments on automatically enriching multimedia resources. Thus, based on the tags already present new knowledge can be derived. As our analysis of tags and queries revealed a second big gap regarding the Opinions/Qualities category for pictures, in [25] we showed that the corresponding task of automatically deriving mood annotations for images based on user generated tags can be approached successfully in a similar manner. Using *Flickr* pictures with the associated tags, we again predicted mood labels according to the model by Shaver et al. [197]. Picture tags were used as features for training a Naïve Bayes Multinomial Classifier with WEKA to predict mood/emotion related tags, which were held out (removed) with the help of the emotion hierarchy and WordNet synsets/synonyms. The results of a 10-fold cross validation were very encouraging, especially for the smaller set of primary emotions. Hit rate @1 ranged from 0.61 to 0.91 (avg. around 0.79); hit rate @3 was close to one for all emotion classes. For those secondary emotions for which we had enough picture instances, also hit rate @1 was often higher than 0.8.

In [21] we also showed that the 'collective intelligence' within the collaborative tagging network *Last.fm*, i.e., the statistical patterns arising from mostly independent user behavior, can be exploited to predict whether a certain song will become a hit. To this end, we trained a Bayesian Network based on artist and track features like average, total and peak number of listeners, initial growth in popularity during the first week, an implicit feature considering a combination of a hub/authority score based on the HITS algorithm [126] and the Billboard peak position of earlier songs of the artist as well as the peak position of the album on which the song appeared.

In the next chapter we turn towards explicitly given social connections between people, the characteristics of weaker and stronger relationships, and the resulting implications for social search.

5 Weak and Strong Ties: Characteristics and Potentials

Especially for young people music is an important part of their life and a central building block of their self-defined identity. Research on the psychology of music shows that music preferences are related to personality traits as well as values. For example, intense and rebellious music was shown to be positively related to openness to new experience [180] while correlating negatively with values of self-enhancement like own success and dominance over others [82]. Struggling between the need for belongingness as well as distinctiveness, adolescents seem to develop their musical taste by deciding on which peer group to belong to. They express their own self-concept by joining a musical community and wearing its 'badge', taking on the stereotypes people have about, for example, 'rockers' (see, e.g. [180]). According to [33], shared musical taste indicates shared values. This perceived similarity in values then leads to social attraction – the missing link in explaining musical bonding.

In this chapter we report experiments on a rich set of factors analyzing online friendship on the social music platform *Last.fm* – a type of social media platform rarely studied so far. We contrast these online links with off-line friendship especially investigating homophily on demographics, taste preferences, and social network overlap. For this, we make use of *Last.fm*'s event calendar indicating physical co-presence of users at concerts. We complement this analysis by automatically predicting both kinds of ties based on the friendship characteristics described.

In the second part of this chapter we will show that consideration of social ties can also inform the automatic prediction of future user behavior in a different type of collaborative Web 2.0 platform: We infer editing behavior on the online encyclopedia *Wikipedia*. As *Wikipedia* does not support maintenance of friendship, social ties are formed implicitly by interacting with one another via user talk pages. Thus, we test the transferability of tie strength indicators to other domains and different notions of social ties. Again, we study various metrics capturing network structure, interaction details, and preference similarity. As a result, a few variables turn out to be especially helpful for predicting social links in both systems studied.

5.1 Analyzing and Predicting Friendship Links in Last.fm

Due to the ease of friending, explicit friendship links in online social networks may be spurious, and it is difficult to differentiate close friends from loose acquaintances. However, sometimes such valuable information may be 'hidden'. Here, we will make use of *Last.fm*'s event calendar to contrast these online ties with off-line links of different strength. On *Last.fm* users connect to 'online' friends as usual, but they also indirectly reveal their 'real-life' friends by listing events that they physically co-attended.

Of course, missing event co-attendance may be misleading as some friends may not accurately administer their event calendar or because music or going to concerts together is not what constitutes a particular friendship despite emotional support, shared sports activities, etc. Then again, there may be cases where a strong event co-attendance tie does not correlate perfectly with user-judged friendship strength as, for example, one meets frequently due to shared preferences on taste or locations, but one interacts little. The same can be true when event co-attendance of a group of people is actually a result of one or a few persons only having strong ties with the others. Lacking manual assessments of friendship strength, this proxy is nevertheless very useful to identify most of a user's *Last.fm* friends she/he has a strong connection with in the real-world – hence, indicating time spent together, sharing the same experiences.

Thus, our work complements both prior 'real world' studies and recent research on tie strength in online networks like *Facebook*. The results may be used in applications for friendship recommendation and ranking or for news feed filtering to overcome problems with spurious links in online networks. *Google+* just recently accounted for the importance of differentiating the various ties people form with introducing user maintained social circles (as did *Facebook* with lists).

5.1.1 Dataset

For analyzing homophily and its implications for tie (strength) prediction, we gathered data from the social music portal *Last.fm*, described in Section 2.1. During spring 2010, we collected user information for over 320,000 users in a snowball tech-

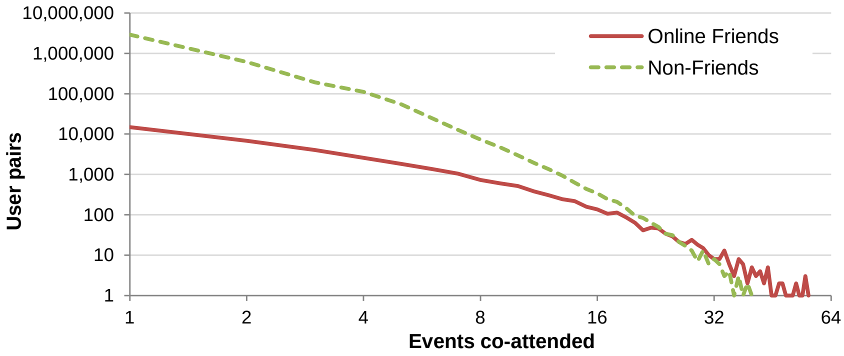

Figure 5.1: Events co-attended by user pairs in *Last.fm*

nique manner using the *Last.fm* API, starting with a single user (the author) and extending the network by friends. For each user, the following details were gathered:

- undirected friendship connections (degree)
- past and upcoming music events (including details like headliner, venue, geographic location, date)
- shouts (messages posted by other users)
- tags (assigned to artists, albums, tracks)
- tracks (recently heard, top songs, positively rated “loved”, negatively rated “banned”)
- albums (top albums and library)
- artists (top albums and charts)

48,527 persons made use of the event calendar, for whom we collected 668,078 user-event-relations to 271,689 distinct events. Co-attendance of all user pairs can be best approximated by a power law like distribution with an exponent of $\approx$ -1.96, which appears as a line on a log-log plot (Figure 5.1). On average, friends attended 1.03 events together, non-friends 0.005.

To calculate the overlap between friendship ties and shared events, we build two different undirected network graphs by taking the 48,527 sample users as nodes and adding a relation between two users if a) they are connected by friendship, b) they are (indirectly) connected since they attended the same event(s). 40,925 edges are in the intersection. Thus, 26% of friendship ties coincide with shared events. To see whether this considerable overlap is significant, we compare to a random graph having the same degree distribution. This null model was constructed by permuting the node labels while keeping incoming/outgoing links. We found 461 edges present in both the random friendship and the event network (0.26% overlap). From the perspective of the event network overlap with friendship is negligible. 1.03% of all event edges coincide with friendship ties.

5.1.2 Analysis of Homophily

To test for similarity among weak and strong *Last.fm* friends, we now analyze agreement on user demographics, taste as well as the local social network. As there are 124,273 friend pairs within our sample of 48,527 users having events listed, we randomly sampled the same number of non friend pairs to compare against.

5.1.2.1 Method

First, we performed a Pearson's Chi-Square test (see, e.g., [140, 79]) for the two categorical variables gender and country. Here, the number of observed pairs with same attribute values (e.g. 'male') and the number of pairs with differing values are contrasted to the frequencies that would be expected assuming random pairing, i.e., based on the probability of each value estimated from the sample. The ratio Observed/Expected (O/E) indicates the assortative strength of an attribute value. It corresponds to the affinity metric for graphs proposed in [164].

For numerical attributes, we calculated distribution means, attribute value autocorrelation, and correlation with strength of the off-line tie[1]. Since both the one-sample Kolmogorov-Smirnov test (see, e.g., [78, 114]) and its adaptation the Lilliefors test [144] showed with significance $p < 0.001$ that none of our attributes is normally distributed, statistical significance of differences on means between friends and non-friends was estimated with the help of nonparametric methods: the two-sample Kolmogorov-Smirnov test (see, e.g., [78, 114]) in case of unequal variances and Mann-Whitney-Wilcoxon (U-Test) [108, 85] in case of equal variances (decided by a Levene's test [136]).

Autocorrelation measures the correlation between the values of a user pair that are not calculated pairwise but are attributes belonging to each user herself/himself. Autocorrelation and correlation with the number of co-attended events are based on Pearson's r, transformed into z-values for assessing significance (see, e.g., [75, 56]). While Pearson is sensitive to the shape of relations and thus may not describe well non-linear relations, there is no intuitive nonparametric equivalent. Thus, we take it as a descriptive indicator useful for comparing the two groups. We note if a

[1] All these descriptive statistics and significance tests for numerical attributes were conducted using the corresponding MATLAB implementations. For details on MATLAB please refer to: `http://www.mathworks.com/products/matlab/`. Documentation on tests provided in the statistics toolbox can be found at `http://www.mathworks.com/help/stats/`

better approximation of the relation can be found through simple transformation, e.g., taking the logarithm to account for exponential or power law distributions.

With our focus on shared musical taste, we additionally perform an experiment tracking down the impact of tie strength on similarities in listening behavior. For this, we adapt the technique used in [57] to predict *Wikipedia* co-authorship. We extend the method to also include a parameter for tie strength:

- Take two snapshots $t1$ and $t2$ of users' weekly charts, two months apart.
- Find all quadruples (u, S, k, ts) where
  - S is a song,
  - u is a user not having listened to the song S at the time of the first snapshot,
  - u had k friends of a given tie strength ts who had listened to the song S at time $t1$.
- $P(k)$ is the portion of these quadruples (u, S, k, ts) given a k and tie strength ts such that u has adopted by having listened to the song S by the time $t2$.

5.1.2.2 Findings

We applied the statistical tests described to different user and user pair attributes relating to demographics, network structure, and taste preferences. The next sections cover in detail the findings for each kind of information.

	Not Friends	**Online Friends**	**Off-line Friends**
Same ***Gender***	0.998	0.955 ***	1.0245 ***
Same ***Country***	1.022 +	8.266 ***	12.571 ***

Table 5.1: Observed over expected frequencies (Chi-Square test) for nominal attributes; $+p$ <0.1, ***p <0.001

Demographic. In Table 5.1 the statistics for the nominal attributes gender and country are shown[2]. For gender, there is a slight, but highly significant, tendency for mixed-gender friendship. We only find 95.5% of the amount of same gender pairs that should be expected given the equal distribution of males (62%) and females (38%) in both populations. However, with increasing tie strength this trend is reversed, now slightly in favor of same gender pairs. For non-friends, the observed frequencies fit

[2]The numbers hold for the entire dataset of 320,000 users.

	Mean			**Correlation Off-line Tie**	
	Online Friends	Not Friends		Online Friends	Not Friends
DEMOGRAPHIC					
Age	23.236	23.295		-0.037 ***	-0.003
Distance (km)	2283.9	4694	***	-0.213 ***	-0.053 ***
STRUCTURAL					
Degree	35.385	33.337	***	–0.048 ***	0.021 ***
Mutual Friends	3.507	0.017	***	0.306 ***	0.072 ***
Mutual Friends Rel.	0.09614	0.0005	***	0.232 ***	0.042 ***
SIMILARITY					
Tag Cosine	0.055	0.010	***	0.051 ***	0.014 ***
Loved Tracks Dice	0.004	0.001	***	0.091 ***	0.026 ***
Top Tracks Dice	0.018	0.001	***	0.074 ***	0.04 ***
Banned Tracks Dice	0.00019	0.00002		0.00003	0.0002
Top Artists Dice	0.13	0.024	***	0.146 ***	0.063 ***
Play count	34497.4	33862		0.001	0.015 ***
Attended Events	14.303	13.657	***	0.401 ***	0.07

Table 5.2: Comparison of means and correlation with off-line tie strength for different user (pair) attributes; ***p <0.001

the expected ones. Not surprisingly, living in the same country calls for high affinity. While for non-friends observed frequencies are almost as expected, it is 8.3 times more likely than chance that the friend of a user is from the same country. As physical co-presence requires geographic proximity, the effect gets stronger for off-line friends connected by events. Table 5.2 and 5.3 show the statistics for numeric attributes. Age shows a right-skewed distribution with a mean of 23.2 in both subsamples of friend pairs and non-friend pairs. Friends are slightly more likely to be the same age. The average age difference is 5.43 while it is 6.61 for non-friends. In contrast to distance, age and age difference do not substantially correlate with tie strength.

As *Last.fm* does not provide user profile fields for location except country, we inferred a user's home by taking the latitude and longitude associated with the town most of the attended events took place. Distance between two people was then calculated as Haversine distance [201]. Of course, the average distance is considerably smaller for friends than for randomly paired users. Pearson's r of -0.213 underestimates the strength of the relationship. After taking the logarithm of both the distance and the number of co-attended events, the correlation coefficient is around -0.52 for friends and -0.17 for non-friends, thus, hinting towards a power law.

	Autocorrelation	
	Online Friends	Not Friends
DEMOGRAPHIC		
Age	0.151 ***	0.008 *
STRUCTURAL		
Degree	0.241 ***	0.019
Mutual Friends Rel.	0.672 ***	0.517 ***
SIMILARITY		
Play count	0.084	0.006 *
Attended Events	0.191 ***	0.007 *

Table 5.3: Autocorrelation along different attributes; $^{*}p<0.05$, $^{***}p<0.001$

Structural. Structural variables give insight into how a user (pair) is integrated within the larger social network. Focusing on simple local metrics, we looked at users' degrees, mutual friends as a measure of link embeddedness, and the ratio of common friends over all friends each user has. The latter ratio is closely related to the number of closed triangles or the neighborhood overlap. It is inversely related to the number of 'forbidden' triads. According to Granovetter [93], the stronger the ties between two pairs in a triad, the less likely it is unbalanced or 'forbidden'. We test this assumption by counting – using Pajek – the frequencies of all four possible types of triads within our online friends sample.

Users in our friend sample seem to be somewhat more active, connecting on average to two more users than users in the random sample do[3]. We find a weak correlation between the degrees of friends, indicating that users somewhat prefer to connect to users that are similar with respect to social activity. There is, however, no significant correlation with the number of common events. The frequency distribution of common friends in our friends sample follows a power law with an exponent of $\approx$ -0.96 (R_2 adj. 0.9948). On average, a pair of friends has about three to four mutual friends, closing 5% of the possible triads to the union of their friends.

For the individual user this amounts to around 9.6% of the friends being mutual. The pairwise difference in the relative number of common friends is also rather small (5.9%). Thus, users usually have as well their 'own' set of friends and are not only connected to a subset of the friends of their friends. This balance is also reflected in the autcorrelation of the mutual friends ratio among friends.

All variables considering common friends show a medium correlation with strength of off-line friendship, with the simplest count being most indicative. Again, a log-

[3]The reported degrees for active users may overestimate average degree in *Last.fm*.

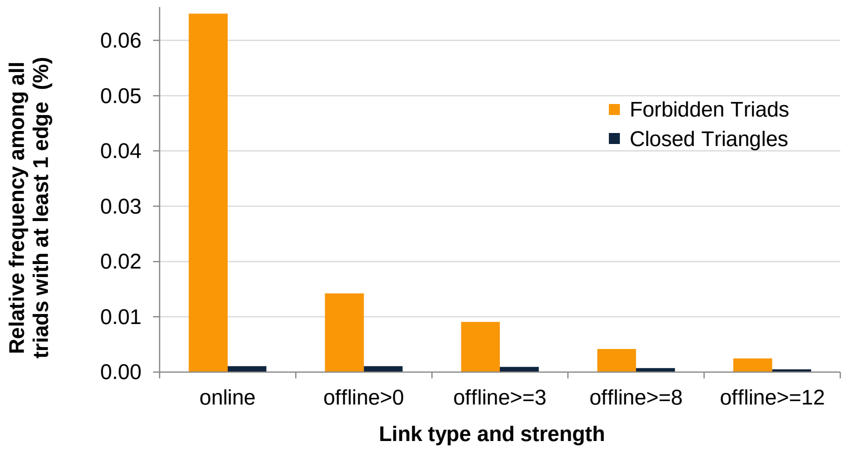

Figure 5.2: Observed closed and forbidden triads as percentage of all triads with at least one edge (online friends sample)

log transformation boosts the correlation for friends (r 0.35). In accordance with Granovetter, the number of forbidden triads decreases with tie strength (r -0.23, see Figure 5.2). However, the online friendship network is already rather balanced with only 0.00003% of all potential triads (0.065% of all triads with one or more edges present) being "forbidden". From the 280,392 forbidden triads in the off-line friendship network (at least one co-attended event) only 11,965 (4%) get closed via a "weak" online link. The ratio becomes even smaller when looking on ties with higher strength only.

Due to the random assignment of users into pairs of non-friends, the corresponding numbers on local structure are close to zero for the comparative sample. The exception is autocorrelation on the relative number of common friends. In the rare cases where a mutual friend exists the ratio is very small but increasing for both users. Only 202 out of 124,273 non-friend pairs have more than one friend in common.

Taste Similarity. Since *Last.fm* and thus user activity is all about music, we now investigate similarity regarding taste in music. We computed the pairwise similarity with respect to top, loved, and banned tracks, tags, and artists relying on the well-known Information Retrieval metrics Jaccard, Dice, and cosine similarity (see [183, 152]). For simplicity, we only report on those metrics that are most indicative for each type of information.

As can be seen from Table 5.2, similarity in general is rather low, still considerably higher for friends than for random user pairs. Correlations of tagging similarity and top or loved track overlap with off-line tie strength are only weak probably due to sparseness in case of tags and loved tracks and due to granularity for tracks. Banned

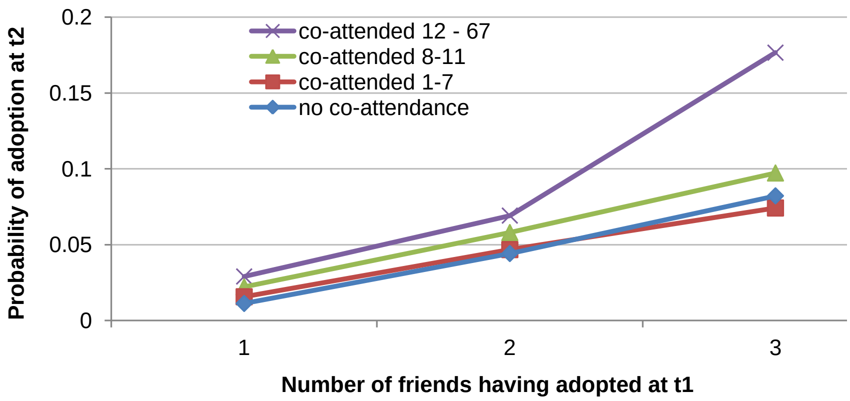

Figure 5.3: Probability of a user listening to a song depending on the number and tie strength of friends having adopted before

tracks are not at all indicative, neither for online friendship nor for tie strength. One reason may be that the feature is not very frequently used: About 50% of the users made such negative ratings (57% in the non-friends sample). Less than 20% of the users banned more than 50 tracks. Indeed, tagging and banned track similarity and correlation increase slightly when only comparing users that used these features. Taking the log of these taste similarity values has a similar (negligible) effect.

Besides, banned tracks are likely bad or 'absurd' recommendations, not representative of taste but the much broader non-taste. The highest similarity values are provided for top artists, correlating mildly with tie strength. As a side note, the cosine metric accounting for user preference in terms of frequency, differentiates less well between friends and non-friends and correlates less with co-attendance. A possible explanation might be that it (over)emphasizes agreement on a few highly weighted common artists.

Figure 5.3 looks at the impact of tie strength on taste similarity from a different perspective – taking time into account[4]. The more close friends listened to a song, the higher the probability that the user will adopt by listening to that song. Whether the reason is social influence or 'pure' prior taste similarity can of course not be answered from this experiment. The result may also be influenced by *Last.fm*'s recommendation system, which could as well exploit such strong off-line ties when generating a personalized radio station based on the social network of the user. As with other partially commercial platforms details of the system's recommendation algorithms are not published.

[4]While for the classes 'no' co-attendance and co-attendance '1-7' the lines follow the same shape for k up to six or higher, our sample data becomes too sparse to make any statements for higher k when looking at strong event co-attendance ties only.

We also compared user activity by looking at the number of played songs and attended events. For the number of played songs no significant difference between the samples nor (auto)correlation is found. With respect to attending a music event, our users in the friendship sample are slightly more active. The correlation with event co-attendance is not surprising as a person's general interest in attending such events influences co-attendance with friends. This correlation is responsible for most of the weak autocorrelation.

Concluding this analysis, we found evidence for affinity regarding location, common friends, and artist overlap as well as for the personal tendency to go to events. However, if we take together all examined variables and look at the partial correlations between each pair of variables, i.e., controlling for the remaining, importance of taste in explaining co-attendance variance diminishes. Artist overlap (r 0.09) is following mutual friends with r 0.233 and distance with r -0.16 as the top ranked indicators of tie strength. In the following we employ machine learning techniques to better understand the interplay of the different variables.

5.1.3 Predicting Ties and Their Strength

Building upon the findings of our analysis, we now predict online friendship as well as the strength of real-world event co-attendance ties to further test the assumption of similarity and overlapping social circles for (strong) ties. Second, we compare our tie (strength) prediction algorithm against user-based Collaborative Filtering (CF, see Section 3.1.2). This is useful to better estimate performance in real applications, and it indicates additional support for the importance of certain user (pair) attributes.

5.1.3.1 Experiments

For our experiments, we use the machine learning library WEKA. Each user pair represents an instance, for which all features are fed into a classifier to decide upon the presence or strength of a tie.

Features. Besides user similarity and local structure, we use a class of tie strength indicators not considered before: interaction between a user pair. Since we are predicting undirected ties of binary *Last.fm* friendship and event co-attendance, we rely only on attributes belonging to the pair of users. Thus, for many numeric attributes, we computed the mean of the two users and the symmetric absolute deviation for each individual user in the pair. The latter should account for similarity, e.g., in case of age, or reciprocity in case of interactional data like shouts to a wall. The next paragraphs give more details about all the features we analyzed.

Demographics and Similarity. The following features capture affinity on demographic attributes and taste similarity. Given the found autocorrelation along online friendship links and the correlation with off-line friendship strength, both kinds of information should help differentiating friends from non-friends and strong from weak friends respectively:

- same and dominant gender
- mean age and deviation
- same country and distance
- mean playcount and deviation
- tagging similarity
- similarity regarding top, loved, and banned tracks
- similarity for top 50 artists

Same gender and country are nominal attributes with the binary values 'yes' or 'no'. If two users have the same gender, the nominal attribute dominant gender names the corresponding gender of the users. Else its value is 'none'. For measuring musical taste overlap, cosine similarity regarding tagging, Dice similarity for top, loved, and banned tracks as well as Dice similarity for the top 50 artists are provided to the classifier.

Structural. To capture local network structure, we provide the classifier with the numeric features below:

- mean degree and deviation
- mutual friends
- mean ratio of mutual friends over degree and its deviation

Our structural features are mean degree and deviation, mutual friends, closed triangles, and the mean ratio of mutual friends over user degree and individual deviation. Degree may be an indicator of exclusiveness or intimacy [87] as it tells with how many other people a user has to share a friend('s attention). Mutual friends and its relative version, the (mean) ratio of mutual friends over user degree, measure the overlap of social circles. Thus, they should be highly indicative of strong ties. Together with its deviation the latter ratio shows the overlap of friends from the perspective of the individual user in the pair. Big deviations hint to unbalance.

Transactional. A lot of tie strength indicators reported in sociology are about interaction data. For the public information available on *Last.fm*, we rely on shouts posted to each others wall:

- minutes since first and last communication
- mean (relative) number of shouts posted and deviation
- mean length of shouts posted and deviation
- polarity of shouts posted

Transactional features used are the minutes passed since first and last public communication via user shouts, mean (relative) number of shouts posted on a friend's wall and its deviation, mean length of shouts posted to a friend and its deviation as well as shout polarity. Time since first and last communication is inferred from the timestamps of these publicly available user shouts. The two variables represent duration and recency respectively (see [87]). Interaction frequency is estimated from the number of shouts exchanged and from its relative counterpart, which moderates the number by the users's posting behavior to all *Last.fm* users. Shout length, on the other hand, points to intensity. Polarity of shouts as a measure of sympathy or affection can be considered a variable corresponding to the dimension of emotional support (see [87]). For inferring how positive, negative, or objective a wall post is, we employed SentiWordnet [70][5], a publicly available lexical resource for sentiment analysis. The words for each English post were analyzed, and positiveness, negativeness, and objectiveness of the shout were averaged over the individual scores for each word.

Algorithms. Algorithm 5.1 summarizes the steps in tie (strength) prediction. For binary prediction of off-line tie existence, all friend pairs having co-attended at least one event make up the positive training examples. From the remaining friend pairs we randomly added the same number as negative examples. Balancing positive and negative examples for each class is helpful to see improvement over a baseline random prediction. To enable more fine-grained assessment of tie strength, in a second run we assigned our friend pairs to one of three bins: no co-attendance, one to nine, and 10 or more events visited together. These classes are based on the logarithmic values of the absolute co-attendance numbers, as inverse of the found power law like distribution (see Section 5.1.1), broken into equally sized intervals of one. The classes represent absent, weak, and strong real-world ties respectively. A tighter definition of weak ties (intervals of 0.5) was found less meaningful.

[5] `http://sentiwordnet.isti.cnr.it/`

Alg. 5.1: Tie (Strength) Prediction

1: Select sample data
For each tie strength class
1a: Retrieve user pairs having co-attended the respective number of events
1b: Randomly sample pairs such that class size equals the size of the smallest class
2: Split sample data set P_{total} into
P_{train} = user pairs used for training the classifier
P_{test} = user pairs used for testing
3: Create features for training and testing the classifier
For each user pair $p_i \in P_{total}$
Compute feature vector $F(p_i)$ with
- nominal attributes for gender and country
- numeric attributes for all other similarity scores, means, and deviations (Section 5.1.3.1)
4: Train classifier on P_{train} using $\{F(p_i); p_i \in P_{train}\}$
5: Make predictions on testset
For each user pair $p_i \in P_{test}$
5a: Compute probability distribution across classes
5b: Assign class with highest probability

However, strict distinction between weak and strong may not be appropriate at all. Actually, tie strength may also be continuous, an issue not resolved so far [87]. Thus, we experimented with a fuzzy variant of the classes above. While the 'no' class is discrete, we allow for overlap of the other two classes like this: 'no', '1-11','8-67'. Technically this is realized by training a multiclass classifier on four distinct classes, but confusions between the overlapping classes are not penalized. Again, class size was balanced. The same procedure was applied for the comparative sample of non-friends. In order to eliminate effects of dataset size, we limited the number of instances per class to those used for friends.

Finally, we infer the binary *Last.fm* online friendship links. Here we randomly sampled around 40,000 friend pairs and around 40,000 non-friend pairs from among our sample of 48,527 users. The algorithm is the same as Algorithm 5.1 with only two classes to be learned, except that we use *Last.fm* friendship as sampling criterion in Step 1, not events co-attended. We used classification via regression on MP5 model trees, i.e., decision trees having linear regression at the leaves [77]. In contrast to other classifiers less suited for numeric and interdependent attributes – such as Naïve Bayes –, Support Vector Machines yielded a comparable performance, but they are computationally much more expensive.

Classes	Instances per class	**Online Friends**	**Not Friends**
‘no’, ‘yes’	39,863	81.57%	91.49%
‘no’, ‘1-11’, ‘8-67’	1,968	80.88%	90.07%
‘no’, ‘1-9’, ‘10-67’	2,926	66.64%	83.64%

Table 5.4: Classification accuracy for inferring strength of off-line ties

5.1.3.2 Results

The proposed machine learning algorithms for online tie and off-line tie strength prediction will now be evaluated with respect to classification performance. We evaluate the learned models by applying stratified 10-fold cross-validation averaging accuracy over runs on all folds. For assessing impact of individual features and feature subsets, we rely on the attribute selection techniques Information Gain and Correlation-based Feature Subset Selection. The latter picks subsets of attributes, in which the single features are highly predictive of the class but do not intercorrelate [100]. It was used together with Best First bi-directional search.

Predicting Strength of Off-line Ties. Table 5.4 shows results for inferring real-world ties and their strength for friends and non-friends. For the easier task of binary off-line tie prediction, we achieve the best results with around 82% accuracy (AUC 0.89) for friends and 91% (AUC 0.97) for non-friends. AUC is the area under the receiver operating characteristic (ROC) curve, plotting the rate of true positives vs. the rate of false positives as the classifier’s discrimination threshold is varied [152]. With respect to tie strength, performance is not impressive given strict classes. Classes that account for fuzzy boundaries between weak and strong off-line ties show good performance, close to the binary task, with around 81% for friends and 90% for non-friends. Thus, the most difficult parts seem deciding about the presence of an off-line tie at all and to draw the exact border.

Based on the Correlation-based Feature Subset Selection the initial feature set could be considerably reduced. Figure 5.4 shows the final feature set for classifying friends and non-friends along the classes ‘no’, ‘1-11’, ‘8-67’. Distance and same country are of course very important, even more so for non-friends. Structural information, simply counted as the number of mutual friends, is also discriminative beyond random for both groups. Thus, even for non-friends knowing that two persons have a common friend increases chances of co-attending an event – thereby closing the ‘forbidden’ online triad.

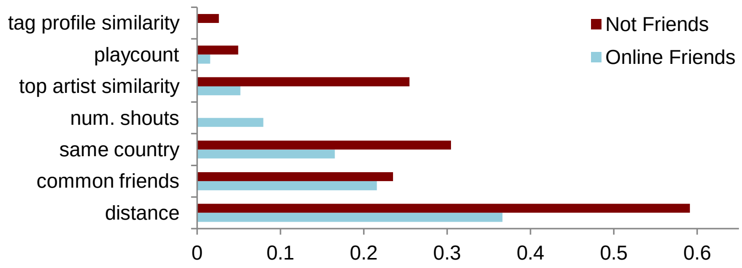

Figure 5.4: Information gain of final features for predicting tie strength (classes: 'no', '1-11', '8-67')

Transactional data is only valuable for inferring tie strength of users that are already online friends. This is expected since communication via shouts is practically absent in our non-friends sample. We had a more detailed look into interaction between our friend pairs (Table 5.5), again by analyzing autocorrelation along the connection and correlation with real-world friendship strength. Given the medium to strong autocorrelations on the (relative) number of shouts, their length, and objectiveness, we find support for mutuality / reciprocity. Significant, yet weak, correlation with off-line friendship strength is just the case for the frequency variables number of shouts and its relative version. Indeed, the classifier only uses the frequency variable number of shouts. In our experiments, sentiment, mutuality, and recency could not add additional information.

For non-friends, shared taste is almost as important as geographic proximity. Thus, frequency of event co-attendance of users not connected online can be well estimated by relying on location, artist taste, and common friends. For friends, taste similarity remains of low discriminative power – even though it is increasing with tie strength.

	Mean	Autocorrelation		Correlation with Off-line Tie	
Days since last Communication	773.08	–		-0.1185	
Days since first Communication	895.36	–		0.0346	
Number of Shouts	1.883	0.5417	***	0.1496	***
Length Shouts	32.51	0.4644	***	0.1061	
Number of Shouts Rel.	0.073	0.3821	***	0.1217	***
Positive Shouts	0.016	0.2886	***	-0.0624	
Negative Shouts	0.010	0.30	***	-0.0259	
Objective Shouts	0.156	0.3720	***	0.0053	+

Table 5.5: Interaction between online friends; $+p<0.1$, $***p<0.001$

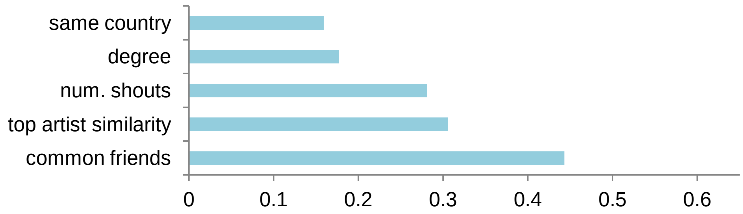

Figure 5.5: Information Gain of final features for predicting binary online friendship

Predicting 'Virtual' Friendship. Predicting online friendship is a much easier task. Given the full feature set, we achieve an almost perfect accuracy of around 94.18%. After applying Correlation-based Feature Subset Selection with subsequent removal of attributes with low Information Gain, we remain with a final feature set (see Figure 5.5) of five, resulting in an accuracy of 93.74%. Again, homophily on country is a strong factor. In contrast to off-line ties, being in very close proximity is not necessary for establishing or maintaining an online link. Taste similarity seems more discriminative for potentially spurious online friendship than for tie strength. Given taste information only, performance is with 77.62% substantially better than random guessing.

Interaction between two users is, of course, an important hint towards online friendship. As it is so discriminative, the simple frequency count seems sufficient for the classifier. The different communication related variables do not significantly complement each other in providing new information. The discriminative power of the few basic structural measures is impressive. Given the number of mutual friends alone yields 83.68% accuracy. Clearly, randomly paired, unconnected users will not have high overlap in friends. However, that 'weak' and 'strong' online friends alike tend to close triads so overwhelmingly often is nice support for network theories.

5.1.3.3 Comparison with Collaborative Filtering

We compare our predictors with a simple standard user-based Collaborative Filtering approach. We calculate similarity between all user pairs relying on the Tanimoto coefficient, as implemented in Mahout Taste[6], with respect to either online friends, events visited, loved tracks, top tracks, or top artists. Tanimoto is an extension of the Jaccard coefficient accounting for overlapping and disjunct preferences by dividing the intersection of two users' lists, e.g., songs both listened to, by the union of both item lists.

[6] `http://mahout.apache.org`

	Spearman's ρ	Kendall's τ
Tie strength prediction	0.7046	0.6606
CF friends	0.6187	0.5485
CF loved tracks	0.5711	0.5467
CF top tracks	0.5422	0.5167
CF top artists	0.5204	0.4568
CF tags	0.3384	0.2810

Table 5.6: Rank correlation of off-line friendship strength with friends ranked based on predicted tie strength or Collaborative Filtering (CF)

Ranking Online Friends based on Off-line Tie Strength. For prediction of friendship strength, we took the remaining 116,340 friendship pairs in our original sample that have not been used during training the classifier. These are mainly friend pairs with none or a few events co-attended, which have not been considered for training due to balancing the distribution of examples per class. For each user pair, tie strength was predicted with the classes 'no', '1-7', '8-11', '12-67'. Collaborative Filtering user similarity was calculated for the pair based on the different types of preference items.

To evaluate the quality of both approaches, we measured the correlation between the resulting rankings and the ground truth ranking of 'real-life' event co-attendance by averaging over those rankings for individual users in the evaluation sample that have a significant correlation ($p < 0.05$). Computing correlations for ranks instead of absolute values is preferable since the classes of our classifier with its associated confidence values do not correspond to the similarity range of the Tanimoto coefficient. We report both measures commonly used for this task: Spearman's rho ρ (see, e.g. [85]) and Kendall's tau τ (c) [123, 1]. Significance is tested via a two-tailed t-test.

Table 5.6 shows that with considering multiple features simultaneously, tie strength prediction better reflects the true ranking of friends according to off-line friendship strength. Both rank correlation metrics show a strong correspondence between ranking based on predicted ties and event co-attendance. For tie strength prediction, the average score remains nearly the same when also taking into account not significant user rankings.

Still, only relying on overlap in friends Collaborative Filtering is pretty strong – emphasizing again the impact of this piece of information. However, with Collaborative Filtering via common friends a lot more user rankings have non significant correlations (the average ρ is 0.3467 not applying the significance threshold). The taste preference indicators show a somewhat lower correlation, with tags and tracks having the least significant individual user rankings. Since this information is so sparse and fine-grained, only few pairs can be assessed in their similarity.

Recommending Online Friends. For comparative evaluation of friendship prediction, we took the remaining 3,598 users from our sample (of 48,527 users with events listed) that were not used in the experiments so far, i.e., they were not part of the data for training the online friendship classifier. For each user in this evaluation sample, we calculated the neighborhood with respect to all other 48,526 users in the sample based on the Tanimoto user similarity coefficient. Similarly, we had the trained classifier decide for all 3,598 evaluation sample users whether they may be friends with each of the 48,526 users.

For evaluating performance, we use success rate at rank k ($s@k$). This metric indicates whether at least one actual friend has been recommended within the top k. In order to enable ranking, we ordered all users predicted as online friends by our approach by classifier confidence. Since for some users there were ties on top confidence values, we averaged success over the best case of having the actual friends before the cut-off point and the worst case having non-friends ranked higher.

Table 5.7 shows that tie prediction significantly ($p < 0.01$, two-tailed t-test) outperforms Collaborative Filtering based on taste-related preference items like events, tracks, and artists. Again, relying on similarity with respect to social connections is the best performing Collaborative Filtering algorithm, achieving best success rates for cut-off values one, three, and five. However, this superior performance over tie prediction is not statistically significant. However, it proves once more the hypothesis of overlapping social circles for friends and thus the value of known common friends for recommender systems. The lower success of tie prediction for small ks is due to the machine learning algorithm predicting for some users many ties on the top confidence values. Starting from around rank seven this effect is negligible, and tie

	$s@1$	$s@3$	$s@5$	$s@10$	$s@20$	$s@50$
Tie prediction	0.33	0.42	0.49	0.60	0.71	0.80
CF friends	0.43	0.50	0.53	0.56	0.59	0.63
CF events	0.30	0.34	0.36	0.39	0.42	0.45
CF top artists	0.27	0.29	0.31	0.34	0.37	0.43
CF top tracks	0.24	0.25	0.26	0.28	0.30	0.34
CF loved tracks	0.23	0.24	0.24	0.25	0.26	0.27
CF tags	0.23	0.23	0.23	0.23	0.24	0.24

Table 5.7: Success rates of friendship prediction and Collaborative Filtering (CF)

prediction significantly improves over all baselines including Collaborative Filtering via friends.

In the following we will investigate whether tie strength indicators can also help predicting co-editing articles in *Wikipedia*.

5.2 Social Ties for Predicting Future Co-editing Behavior on Wikipedia

Now we want to see how social interaction and other traces users leave in digital records on the web can be exploited to predict future behavior, here in form of article edits in the online encyclopedia *Wikipedia*. More precisely, we ask: Given a social interaction network of users in *Wikipedia*, can we infer co-editing patterns for user pairs, i.e., that one user in the pair will adopt behavior by editing the same article as her/his "friend" did before? If so, which attributes of the user pair or the two individual users are most indicative?

As *Wikipedia* does not explicitly support friendship, a social link between two users is inferred if one person edited the user talk page of another user. This work was partially inspired by [57], where authorship on *Wikipedia* articles is predicted based on social links, i.e., depending on how many friends edited an article before. As the goal was to differentiate effects of social influence from similarity, the predictions were compared to predictions based on user similarity and were found superior. Here, we study a rich set of factors trying to predict *Wikipedia* co-authorship. The results can be used, e.g., in recommender systems suggesting articles to edit like SuggestBot [54][7].

5.2.1 Dataset and Preprocessing

Wikipedia is a free online encyclopedia crafted by collective effort of web users. Everyone can edit articles anonymously or via their user account. A few power users and elected administrators do most of the work though. It is meanwhile available in over 270 languages. As of August 2012 the English *Wikipedia* has 4,028,130 articles. A lot of research has investigated motivations for contributing to the *Wikipedia* [172], studied conflicts, edit wars, and vandalism [125, 217, 218], or compared its quality to traditional expert authored encyclopedias like the Britannica [88], etc.

For our experiments on inferring future co-editing behavior based on social interaction, we used the full *Wikipedia* dump from January 2008, containing all articles

[7] `http://en.wikipedia.org/wiki/User:SuggestBot`

with all their revisions[8]. Dealing with this data is not trivial as it is about 18 GB in highly compressed 7zip[9] format. From this dump we extracted on the fly all article and user talk page edits, including page id, revision, author, and timestamp.

In order to build the social network, we draw a directed edge (or arc) from each user leaving a post on another user's talk page to the owner of that page. The two users in such a pair are henceforth called contributor and page owner respectively. We deliberately ignore here anonymous users, having no username/id, but only an IP address. Also, in a few cases there was a need to do some minor data cleaning activities like merging multiple accounts. For each user post, we store the page owner, contributor/author, timestamp, revision id, whether it was "minor" as well as the new plain text of the post. Each revision of the user page gives a snapshot of the page at this particular time, combining text snippets from diverse authors with that of the current author. To identify which text insertions or deletions can be attributed to a user, we employed the Diff Match and Patch libraries[10], which implement Myer's diff algorithm [167]. If given for a post, we also analyzed comments. Such comments are usually short summaries of what has been changed in this revision.

Our *Wikipedia* social network of ties, representing user talk page edits, consists of all such interactions up to 2006-12-01, the time of our first snapshot $t1$. By this time, there are somewhat over 413 thousand distinct users in the sample. Around 402,500 of them have their own talk page with edits. Interestingly, we find only about 71 thousand contributors actively posting on other users' pages, thus, being responsible for the over 2.3 million single interactions. These correspond to 1,175,555 directed edges, only 177,975 (around 15%) of which are reciprocated, i.e. bidirectional. Thus, the number of distinct user pairs is 997,580.

Everything after $t1$ is considered the future behavior of this set of users. For computing the number of future co-edits, we apply the approach presented in [57, 12] and used before in Section 5.1.2 when analyzing adoption in listening behavior: We count the number of pages each user in the pair edited after $t1$, such that the page has been edited before $t1$ by the friend but not by the user herself/himself. Since we have directed relationships, we thus capture adopted behavior pointing from the page owner to the contributor and vice versa. We now describe the experiments for predicting such future co-editing.

[8]available at: `http://snap.stanford.edu/data/wiki-meta.html`
[9]`http://www.7-zip.org/7z.html`
[10]`http://code.google.com/p/google-diff-match-patch/`

5.2.2 Predicting Future Edits

As in our work on friendship in *Last.fm*, we conduct some basic statistical analysis complemented by employing machine learning techniques offered within WEKA for our co-editing predictions. Here, each directed edge in the social network represents an instance having various attributes or features.

5.2.2.1 Features

Since *Wikipedia* is about collectively building an online knowledge base, articles are in the focus, not users. As such, *Wikipedia* has user and user talk pages that can be freely edited by the user or others, but it does not have (semi)structured user profiles including, e.g., demographic user information. Thus, we resort here to the information available: structural information about the social network topology as well as detailed characteristics of underlying user interactions. We also measure user similarity in terms of overlap on edited articles. For those attributes that are not calculated pairwise but do belong to the individual user within each pair, we always provide the corresponding feature for both page owner and contributor. This is important because our network is directed, and we predict directed adoption in future behavior. For brevity, we do not differentiate these in the following description.

Structural. Attributes capturing even only simple local network structure have been shown to be highly indicative of, for example, friendship strength. Consequently, we also include topological features here. For each user in a pair, we calculated[11]:

- indegree
- outdegree
- number of common friends as common contributors
- Jaccard overlap of common friends as common contributors
- number of common friends as common page owners posted to
- Jaccard overlap of common friends as common page owners posted to
- betweenness centrality
- closeness centrality
- eccentricity
- Eigenvector centrality
- PageRank

[11]Most of these are standard social network metrics, see, e.g. [222].

- clustering coefficient
- strongly connected component
- weakly connected component

Indegree and outdegree tell how many users in the *Wikipedia* social network have written on the current user's page and how many users this user wrote to herself/himself respectively. Similarly, common friends as common contributors gives the number of other *Wikipedia* users who posted both on the page owner's and the contributor's talk page. For common friends as page owners, both users in the pair under consideration contributed to these other users. We also take the Jaccard indexes [152] dividing the common friends by the union of friends each individual user has.

Since we have a complete social network and not "only" a part of it (as was the case with our *Last.fm* crawl), we also include features going beyond the pure local ego-networks. The remaining attributes are all calculated by using well-known social network analysis metrics as implemented in Gephi[12]. Betweenness centrality of a node counts how often shortest paths between any two nodes in the network pass through the node, i.e. the user. Closeness centrality, on the other hand, measures the user's average distance to the other nodes in the network. Eccentricity is the largest shortest path (i.e. the largest geodesic distance) to any other node. Eigenvector centrality is a metric capturing importance or influence of a node within the network based on its connections to other important nodes. The PageRank [175] implemented within Google's search engine is a variant of this measure.

The clustering coefficient of a node measures how well its neighbors are connected among themselves. A local clustering coefficient of one would mean complete connectedness among a node's connected nodes. Finally, we detect strongly and weakly connected components and provide the respective component ids as classification features. A strongly connected component is a set of nodes that are all reachable by each other via an existing path. For weakly connected components, link direction is ignored.

Transactional. With the transactional features we aim at capturing recency, duration, frequency, intensity as well as sentiment of user interaction:

- minutes since first and last communication (overall)
- number of posts sent to each other

[12]for details see: `http://wiki.gephi.org/index.php/Category:Measure`

- (average) number of words added and removed (overall)
- average polarity of posts (overall)
- smilies in posts,
- number of posts marked as "minor"
- average number of terms for comments accompanying posts
- smilies in post comments
- polarity of post comments

First, we note down the minutes since the first and last communication initiated by the page owner to the contributor and vice versa. We also include the corresponding times for communication with whatever user within the network. For frequency, we provide the number of times the user talk page has been edited by the other user. Pointing towards intensity, we analyze the amount of words added and removed in posts to each other's talk page as well as the average intensity for each user with respect to all users in the network. For determining polarity of user page edits and their attached comments, we again made use of SentiWordnet. Regarding smilies, we parsed the added text for occurrences of smilies as listed on the "List of emoticons" page in *Wikipedia*[13]. Besides the frequencies for the individual types like "happy", we also include the total number of emoticons used and, again, an average for both users showing overall tendency for smiley/emoticon usage. A revision can be marked as "minor"; we count the number of corresponding posts to each other. Finally, if present, positivity and negativity scores as well as frequencies for the different types of smilies are given for comments.

Similarity. For comparison, we are also interested in seeing how predictive similarities on articles edited in the past are for future co-editing behavior. For this, we computed the *weighted Jaccard coefficient* as in [57]:

$$Jacc(\vec{c}, \vec{p}) = \frac{\sum_{j=1}^{n} \zeta_j \min(c_j, p_j)}{\sum_{j=1}^{n} \zeta_j \max(c_j, p_j)}$$

where $\vec{c}$ and $\vec{p}$ are the pages edited by the contributor or page owner respectively and ζ_j is a weighting factor inversely proportional to the number of users having edited the page.

5.2.2.2 Experiments

We randomly sampled 120,000 edges from the *Wikipedia* talk network. First, we will provide some basic descriptive statistics like means, autocorrelation, and correlation

[13] `http://en.wikipedia.org/wiki/List_of_emoticons`

with absolute co-editing numbers. In this analysis we removed outliers by deleting sample pairs having co-edit values occurring less than 10 times in the entire sample. Co-edit values of more than 45 were that rare in our sample and are thus excluded. Lilliefors test showed again that none of our variable comes from a distribution in the normal family. When assessing significance of differences in means, thus, we again rely on the two-sample Kolmogorov-Smirnov test and the Mann-Whitney-Wilcoxon (U-Test).

Then, in order to explore whether and which combination of the many features described are useful for predicting edits on an article the friend worked on before, we convert the numeric class value of future co-edits into a binary nominal variable {'yes','no'}. To better see emerging patterns, class size is balanced. The final numbers are each 19,104 negative and positive examples for inferring co-editing behavior at $t2$ on the side of the contributor. The respective number is 18,751 for runs predicting later co-edits performed by the page owner.

Here, we used bagged decision trees [35] based on REP trees[14]. Classification via regression on MP5 decision trees had comparable results. Other classifiers, including even sophisticated ones suitable for numeric and interdependent attributes like Support Vector Machines, performed less well. As before, classification performance is measured by applying stratified 10-fold cross-validation averaging accuracy over runs on all folds. The attribute selection techniques of Information Gain and Correlation-based Feature Subset Selection are exploited to help reducing the large feature set to the really indicative ones.

5.2.3 Results

Before reporting results of our machine learning experiments, we provide descriptive statistics on the structural, interactional, and similarity attributes used.

5.2.3.1 Descriptive Statistics

The frequency distributions of co-editing patterns are very similar no matter the direction. Both follow a power law function with low values being highly frequent and high numbers of co-edits occurring very rarely as the long tail. The exponents of the functions are -2.065 for co-edits at $t2$ by the contributor and -2.058 for future co-edits by the page owner. The mean number of future edits is 0.648 (SD 2.799, median 0) for the contributor and 0.62 (SD 2.675, median 0) for the page owner. We find quite some reciprocity: 83.26% of the sample edges have either both numbers

[14] `http://weka.sourceforge.net/doc/weka/classifiers/trees/REPTree.html`

	Mean		**Autocorr.**	**Corr.** $t2$ **P**		**Corr.** $t2$ **C**	
	P	C		P	C	P	C
STRUCTURAL							
Indegree	100.98	189.74	-0.0879	0.1503	0.0801	0.2008	0.0501
Outdegree	185.10	754.14	-0.1304	0.1671	-0.0175	0.1653	-0.0179
Betweenness	20,858,885	65,512,573	-0.1112	0.151	0.0096	0.1684	0.005(ns)
Closeness	2.0296	3.1654	0.2026	0.1407	-0.0764	0.1391	-0.0614
Eccentricity	4.3425	6.9559	0.1904	0.1554	-0.0453	0.16	-0.0512
Eigenvector	0.0701	0.1231	-0.0258	0.1463	0.0887	0.2057	0.05
PageRank	0.000053	0.000099	-0.0753	0.1104	0.0618	0.1574	0.0313
Clustering coeff.	0.0681	0.0403	-0.0002(ns)	-0.0423	-0.0197	-0.0437	-0.0181
Common IN	5.3667		–	0.2259		0.2417	
Common OUT	5.4051		–	0.242		0.2518	
Jaccard IN	0.0135		–	0.1674		0.1702	
Jaccard OUT	0.0086		–	0.1191		0.121	
INTERACTION							
No. of posts sent	0.9426	1.9492	0.5256	0.1084	0.0686	0.1025	0.08
Positivity posts	0.0157	0.0477	0.126	0.1127	0.0265	0.1029	0.0205
Negativity posts	0.0144	0.043	0.1248	0.1028	0.0178	0.0964	0.0176
SIMILARITY							
W. Jaccard coeff.	0.0015		–	0.0276		0.0224	

Table 5.8: Means and autocorrelation along different attributes as well as correlation with co-editing behavior through the page owner (P) or the contributor (C)

on future co-edits equal to zero (75.86%) or both higher than zero (7.40%). There is also a mild correlation of r 0.321 between the absolute numbers for a pair.

Table 5.8 shows the statistics for all structural variables, for similarity, and for selected attributes capturing pairwise interaction. If not stated otherwise (i.e., as "ns"), all correlations as well as the differences in means between values for page owner and contributor are highly significant ($p < 0.001$).

Structural. For the *Wikipedia* user talk page social network as a whole, we can report a maximal distance between any pair of nodes (i.e. diameter) of 12.0, an average path length of 49.32, an edge completeness or density of 0.000007, and an average clustering coefficient of 0.048. Our social network comprises one large strongly connected component of 57,520 users and 325 strongly connected components of size two to seven. The remainder out of the 355,645 strongly connected components are made up by isolated users. A little more than half of the user pairs are both part of the same big strongly connected component. Pairwise correlations between variables reveal some redundancies. For example, closeness and eccentricity correlate almost perfectly, with eccentricity better indicating future co-edits. In our network the user's largest shortest path is usually around 2.2 times its average distance to other nodes. Similarly, indegree correlates strongly with Eigenvector centrality and PageRank (Pearson's r 0.96 and 0.93 respectively). This is not surprising since both latter metrics operate on incoming links.

Users contributing to other users' talk pages have considerably higher values for indegree (and thus for Eigenvector and PageRank), outdegree, betweenness, closeness, and eccentricity. Thus, contributors are socially more active, writing more often on others' talk pages but also receiving more posts by others. As such they are much better connected appearing more often on shortest paths between other nodes and having a higher average distance to others (and, thus, a higher largest geodesic). The less active page owners, in contrast, maintain fewer connections, which are better connected amongst each other as is indicated by the higher clustering coefficient. The low values for autocorrelation in the pair give a similar picture of dissimilarity in our sample pairs. Only for closeness/eccentricity there is minor correlation along the edges. The number of common friends for a pair is with 5.4 comparable for both outlinks, i.e., people both wrote to, and inlinks, i.e., people having written on both users talk pages.

Looking at how indicative the different structural variables are regarding adoption of behavior, we see that both kinds of common friends are best correlated with the number of future co-edits – with common outlinks being a little more correlated with future adoption by the page owner and (even a little more so) with adoption at $t2$ through the contributor.

For all other attributes, it is interesting to note that mainly the attributes of the page owner are relevant regardless of whether we want to predict co-editing at $t2$ by the page owner or the contributor. We suppose the reason is that the set of distinct contributors is considerably smaller. In our sample the number of distinct page owners is more than three times the one for contributors. Thus, attribute values are more homogeneously distributed in the contributor group as reflected by smaller standard deviations. Page owners, on the other hand, have extremely biased distributions, for example, often having a value of zero for outdegree and one for indegree.

High values on these structural user variables imply better connectedness, i.e., more (social) activity. This page owner activity probably is the cue for predicting future co-edits. In the case where the contributor adopts by editing a page that was edited before by the user whose talk page (s)he wrote on, the page owner's structural characteristics may hint towards her/him being influential or, on the contrary, controversial (both could be coupled with social activity). As contributor attributes only indegree/PageRank/Eigenvector, potentially indicating influence, show a minimal correlation with the probability of the page owner adopting.

Interaction. Not surprisingly, contributors have higher values for the number of posts sent to the other user. An autocorrelation of 0.52 shows that there is quite some tendency for reciprocity and balance. However, this effect comes mainly from the many zero co-edits on both sides and from the around 35 thousand pairs with mutual communication. Restricting the edge set to those with bi-directional edges autocorrelation is comparable with 0.54, but differences between page owner and contributor disappear regarding the mean number of posts sent to the other user, the amount of words added, positivity and negativity scores. There are also no substantial differences anymore regarding words removed and all corresponding values for the overall behavior of page owners and contributors. Except for the number of posts exchanged, autocorrelation is very low or actually absent.

Among all smilies, "happy" and "laughing" are most often used, followed by "wink". The corresponding values are considerably higher for links with mutual communication, showing minor autocorrelation on usage of "happy" (r 0.10). Other types of smilies are hardly ever used. In our machine learning experiments we thus remove such 'useless' smilies used by less than 100 distinct users.

Correlation with the number of absolute co-edits is again higher for variables concerning interactions initiated by the page owner though they are still very small. R is around 0.1 for the number of posts sent when measuring on all edges and around 0.075 when only considering bi-directional ones. For the contributor the values are 0.08 and 0.07 respectively. Obviously, these features do not add a lot of valuable information beyond indicating the fact of communication, which is already implied by the very existence of the social link and thus true for all sample pairs. The same applies to comments. Here, the number of comments by the page owner is redundant with the number of posts sent.

The most useful interactional feature is time since last communication initiated by the page owner. It correlates with r -0.141 with adoption at $t2$ by the page owner and with r -0.127 with contributor adoption (-0.093 and -0.085 respectively for mutual links only). However, recency of page owner communication overall is an even better indicator with r -0.181 for future edits by the page owner and r -0.184 by the contributor (-0.144 and -0.106 for bi-directional edges). One likely explanation may be the increased number of values available when regarding recency of communication within *Wikipedia* in general – hinting again to the importance of capturing activity. Here, recency and duration as time since last and first communication are redundant as they correlate almost perfectly with each other for both cases.

Similarity. Intuitively one will consider user similarity with respect to articles edited important for predicting future co-edits on articles. Given the size of the *Wikipedia*, mean article overlap is naturally small with 0.0015 (SD 0.012, median 0.00005). Correlation with the absolute numbers of co-edits is negligible no matter which direction is concerned. However, in the following machine learning experiments, capturing the interplay of variables, we will see that in combination with other attributes there is an indicative value.

5.2.3.2 Automatic Classification

We achieve a best classification performance using the full feature set described above excluding, however, the page owner's and the contributor's membership in strongly and weakly connected components. The accuracy is 82.01% for predicting adoption in terms of co-editing behavior at $t2$ by the page owner and 83.39% for future adoption through the contributor. Interestingly, the strongly connected component feature is highly misleading in both cases as it leads to over-fitting the bagged decision tree on the training data. As a result, over 90% accuracy are achieved on the training data but only 74.09% and 75.02% respectively when doing cross-validation.

For predicting page owner behavior, the top 10 ranked attributes with respect to Information Gain are: time since last communication by the page owner overall, the (deceptive) strongly connected component (s)he belongs to, her/his indegree, betweenness, outdegree, closeness, PageRank, clustering coefficient, eccentricity, and common friends of the pair, both have written to. Most useless are all individual types of smilies as well as last and first communication initiated by the contributor. When aiming at predicting later edits by the contributor, the picture is similar. There are minor differences in that Eigenvector centrality comes into the top 10 attributes replacing common friends out and a changed order with time since last communication by the page owner overall falling to place 10.

As some of these attributes correlate highly with each other and, thus, do not add informational value, we again used the Correlation-based Feature Subset Selection to reduce the feature space (with bi-directional Best First search). Figure 5.6 and Figure 5.7 show the final feature set for predicting co-edits by the page owner and

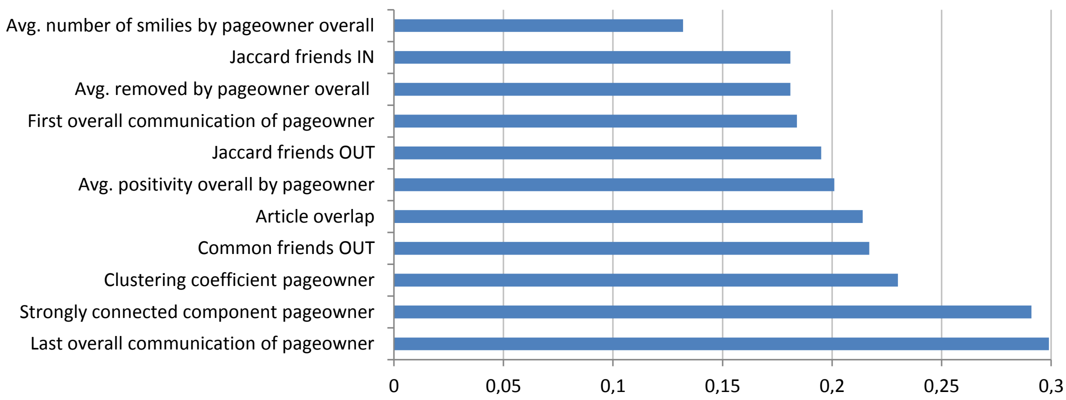

Figure 5.6: Information Gain of final features for predicting future co-edits by the page owner

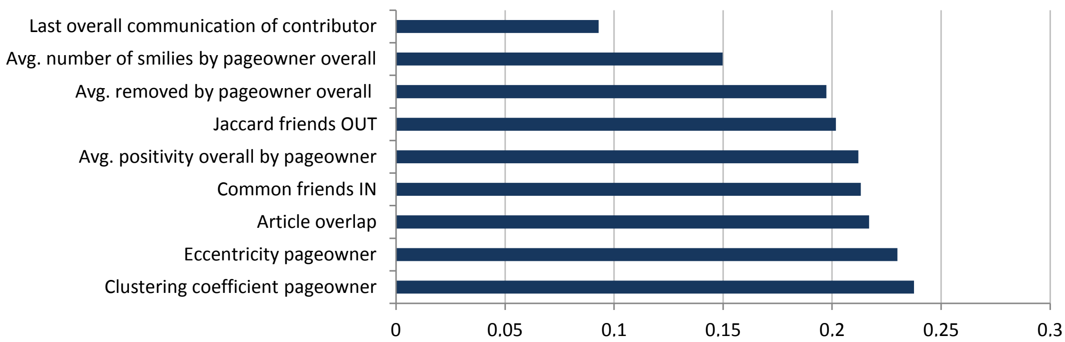

Figure 5.7: Information Gain of final features for predicting future co-edits by the contributor

the contributor. With these sets we still achieve 81.24% (after removing strongly connected components, else 73.31%) and 82.25% respectively.

Common predictive attributes are: article overlap, average positivity, average words removed, average emoticon usage of the page owner, her/his clustering coefficient as well as common friends both users contacted measured via weighted Jaccard. For adoption by the contributor, the simple number of common friends as inlinks is indicative while it is common outlinks for page owner adoption. Thus, the indicative value of (local) network structure is again striking. Relying on such network topological features alone yields around 79% for both directions compared to 72% achieved by the 'default' approach of considering article overlap only.

As already seen in the statistics section, details on pairwise interaction do not help a lot for the task at hand. Overall user activity is better suited. Time since last overall communication of a user as an activity index is useful to consider when predicting her/his tendency to engage in a co-editing activity.

It is important to note again that user interaction is per definition already included by the social ties drawn – lacking explicit friendship links. Second, the found patterns in adoption behavior do not necessarily imply that the reason is social influence. The causality aspect can not be answered by our experiments. Future work may employ techniques like matched sampling (see, e.g. [7]) to differentiate causes, for example, authority and reputation.

5.3 Discussion

In this chapter we first analyzed friendship characteristics in the music platform *Last.fm* and showed that automatic prediction of online as well as off-line ties is feasible. In our study on *Last.fm* we found evidence for affinity regarding location, common friends, and artist overlap as well as for the personal tendency to go to events.

Our findings are in line with the results on demographics and degree reported in [210] for *Myspace*. Regarding taste the slight homophily bias found here is not consistent with the tendency of users differentiating themselves from their friends in *Myspace* [148]. However, *Myspace* is broader in the taste items treated. It is more a general interest networking site than *Last.fm* with its focus on music. Also, explicitly "crafted" statements on user profiles may deviate from implicit taste profiles based on songs listened to. While in a user questionnaire on *Last.fm* [16] homophily among online friends was found for age and shared taste only, here we clearly see a tendency to bond with people from the same country. The difference may be due to that study randomly comparing to one friend only. The tendency increases with tie strength, which is again consistent with [16] regarding geographic proximity.

The study reports that taste similarity does not correlate with relational development, i.e. tie strength. Here, we find shared musical taste to be increasing with the strength of the event co-attendance off-line tie. However, while the two notions of friendship (strength) may not be comparable, we also find that the additional predictive value of this shared taste is indeed negligible. This hints to the importance of other, probably more social aspects, beyond taste preferences, making it harder to distinguish weak and strong ties for real-world friends. It would be interesting to see whether having data about private communication or communication across other media could explain the remaining variance. Yet, taste similarity is a feature worth considering when predicting whether there is an online link at all.

The bad performance of interaction data solely may be due to sparseness and attributed to the nature of *Last.fm* and other taste networks with a specific purpose. While there is usually no transactional information in case of non-friends, even friends do not necessarily communicate publicly via walls nor is communication exclusive to *Last.fm*. Having data about private interactions or more social facilities like 'liking', etc. may yield better results. For the public communication that was observable, recency was found to correlate weakly with tie strength. The impact of duration of the *Last.fm* friendship could not be verified [16] when operationalized as time since first online communication. For all tasks, frequency alone was sufficient. Though [156] concluded that frequency of communication can be misleading, in online social networks like *Last.fm* it may collide with intensity as one more deliberately communicates and not justs happens to meet each other.

However, especially for the task of predicting online friendship, structural and similarity information alone are very valuable. The predictive ability of the simple local measure of common friends is impressive. For non-friends, this indirect connection via a common friend may carry some similarity information, e.g., location or taste, or two users actually know each other in real life, but they forget or decide not to connect online. This impact agrees with prior findings on the importance of even simple local network topology [91, 198, 142]. As we could see and as suggested by Granovetter [93], this tendency gets stronger for close friends. Together with the found balance, which is increasing with tie strength, this supports sociological theories like the homophily and weak tie/forbidden triad assumption.

Our additional experiments on predicting future adoption behavior in *Wikipedia*, i.e., co-editing an article edited before by a 'friend', as well proved the value of such simple metrics capturing local network structure. Together with the clustering coefficient, indicating how well a user's neighbors are connected among themselves, the raw number of mutual friends was one of the most indicative features for the classifier: common friends both page owner and contributor had written to in case of predicting future adoption by the page owner and common friends who had written on the page owners talk page as well as the contributor's talk page in case of inferring adoption through the contributor. These local network features were superior to (almost all) global metrics considering information about the network as whole. Building a classifier on the network topological features alone, we achieve around 79% for both directions – just 3-4% less than the best run.

Of similar importance is the article overlap attribute, ranked high according to its Information Gain. Relying on only this feature yields 72%, which is still substantially better than random guessing. As in the case of our *Last.fm* experiments, detailed information on interaction between a user pair, e.g., intensity measured via the number of words exchanged, is of little value. The pure fact of interaction seems almost sufficient. For our *Wikipedia* experiments such interaction is given per definition, implied by the notion of a social tie. Like for friendship strength prediction in *Last.fm*, recency of communication is one of the more indicative interactional attributes. Though, recency of communication in general, i.e., writing to any user within *Wikipedia*, is highly correlated and an even better indicator, probably indicating (social) user activity.

Thus, the approach of using tie strength indicators – automatically learning their weights – has been shown to be useful for different domains with different notions of tie strength. Second, (the importance of) some variables seem to generalize. Since *Wikipedia* does not have user profile pages with structured fields for demographics, unfortunately we are not able to further test the homophily assumption and, thus, transferability of the corresponding tie strength indicators, e.g., for country of origin, geographic distance, gender, etc.

6 Conclusion and Outlook

Online social networks offer great data to analyze and experiment with for enhancing user profiling, search, and recommendation. The concept of tie strength seems a promising framework for identifying the diverse potentials different online social relations can bring. First, collaborative tagging provides reliable, non-redundant, and interpersonally valuable metadata, which can be used to enhance retrievability of resources as well as to estimate user-user or user-item similarities. For this, no explicit friendship relationships need to be given.

The results of our comparative tagging analysis provide more insights into the usefulness of different kinds of tags for improving search. Unlike most earlier work, here we investigate questions regarding user tagging behavior contrasting diverse popular tagging systems: *Delicious* for general web pages, *Flickr* for pictures, and *Last.fm* for music resources. The comparative study extends prior work on tagging systems by establishing a simple, yet comprehensive tag classification scheme applicable to different systems with different kinds of resources. Moreover, we report descriptive statistics of tag type prevalence for each system in general, but also by differentiating very popular, somewhat frequent, and highly idiosyncratic tags.

We find that social annotations are to a large extent factual in nature. However, participants in our additional questionnaire on tagging also valued subjective tags for retrieval, assuming quite some overlap in judgment. Other studies have shown that users turn to their social networks in particular with subjective queries. Focusing on the value of user generated tags for search and recommendation, we contrast the patterns found for tags with queries from the AOL search engine log, assessing types of queries, types of resources asked for as well as query popularity. In our query log analysis we indeed find a considerable amount of subjective queries, asking, e.g., for context or opinions.

As these are partially underrepresented as textual tag labels, we performed classification experiments for inferring music moods (as opinions) and themes (as usage context) – thereby enriching music tracks with additional information often queried for by users. The algorithms proposed rely either on user tags, on audio features, or on combinations of both. In contrast to earlier approaches exploring the separate use of *Last.fm*, audio content, Web documents, surveys, or annotation games to predict (the likelihood) of all kinds of tags, our work explicitly focuses on inferring mood and theme annotations. While previous attempts to associate mood labels to music songs often rely on lyrics or audio information for clustering or classifying song corporas, our algorithms exploit both audio information and social annotations from *Last.fm*. Thus, we complement work on automatically classifying music mood/emotion based on audio features by using *Last.fm*'s valuable folksonomy information for inferring moods and themes. Whereas in our earlier experiments only tags were used for deriving moods, themes, and styles/genres, here we also investigate fusion with audio-based methods.

Our results show that both sources provide helpful complementary information and should be merged in order to achieve improved classification performance. Using our algorithm music becomes searchable by associated themes and moods by posting textual, descriptive queries. For future work, some of the promising ideas to be further investigated refer to refinements of the moods and themes clusters, as well as to other possible combinations of the audio-based and tag-based classifiers, i.e. metaclassifiers.

Regarding explicitly given ties, we focused on characterizing and identifying weaker and stronger ties. As making friends online is effortless, friendship links are potentially spurious. People are said to bond with people alike, thus close contacts are supposedly very similar to oneself. Here, we study two notions of friendship on the social music network *Last.fm* – a type of platform rarely studied. Based on a rich set of factors extracted from the digital records on *Last.fm*, we characterize online and off-line ties focusing on homophily in particular, and we learn to predict both kinds of ties and their strength. Relying on demographic information like location, simple local network structures, interaction frequency as well as similarity in taste, we can distinguish different levels of event co-attendance, i.e. off-line tie strength. We find support for similarity along social ties and, in particular, strong overlap in social circles increasing with tie strength. The comparative tasks of predicting event co-attendance for non-friends as well as of predicting online friendship are considerably easier as here taste is more discriminative.

Thus, we complement prior work on general purpose online networks like *Facebook* and *Twitter* by transferring the problem of tie (strength) prediction to the taste domain. Here, preference or interest similarity is an important ingredient and thus is to be considered in much more detail. In contrast to earlier work, we do not rely on explicit user ratings or statements as found in questionnaires or on user crafted profiles. Instead, we exploit a variety of implicit, observable user preference indicators like tracks listened to, favorite artists, tags used, etc. Similarly, we do not require users to manually rate tie strength, but we make use of an interesting proxy *Last.fm* offers: physical co-attendance at events listed in the event calendar.

With our experiments on predicting future edits on *Wikipedia* based on adoption behavior of a friend, with social ties defined as user interactions on user talk pages, we showed the transferability of automatically learning the importance of tie strength indicators to other Web 2.0 platforms, belonging to different domains and having different notions of social ties. We augment earlier work on predicting future co-edits in *Wikipedia* by taking into account a variety of typical tie strength indicators like communication frequency or recency as well as network metrics like clustering coefficient, betweenness centrality, etc. Our results indicate that a few variables – in particular simple network topological metrics – turn out to be especially helpful for predicting social links in both systems studied.

In future work, we plan to incorporate tie strength into recommender systems to tackle problems like lack of novelty or diversity, thus, testing Granovetter's hypothesis on the strength of weak ties. Few work has been done so far on how to incorporate tie strength into information retrieval and recommendation systems. The answers to these questions have direct implications on personalization approaches for information and people search, filtering, and ranking. Depending on the task at hand the potential of strong or weak ties can be exploited. While with sensible information, for example, strong ties will be trusted more, diversity and serendipitous encounters can be enforced via incorporation of weak ties. To raise awareness, e.g., in news feeds or visualization, prominent ranking of strong friends may be aimed at.

Future research should also extend work on modeling ties based on explicit or implicit (e.g. tags) indicators. For different domains, system designs, and available transactional data, results may deviate from the previous studies – especially regarding homophily. Characterizing the relationships people form online and studying how these relations (or their attributes like taste profile similarity) evolve over time are important to assess the value of ties for improving search and recommendations. Applying standard social network analysis procedures on the new, large datasets will also shed additional light on larger community structures around strong and weak ties in general. A lot of other questions can be raised, for example, how – considering different tie types and strength – the composition of the global network as compared to local ego-networks correlates with different kinds of behavior. Differentiating effects and causes of correlated behavior is a further important issue.

Last but not least, exploring tie strength and its potentials in specific target scenarios, for example, resource-based learning as in the CROKODIL project is interesting. From the case study on usage of social networks and collaborative tagging, we could see that social networks are an important channel for both private and learning oriented communication. Complementing structural and content-based approaches, e.g. on tags, for personalized recommendations, we plan to experiment with knowledge propagation along the social network of CROKODIL users to verify the potential of different tie types and strengths. Assessment of the strength of ties in such learner networks is the first step. Interactions between CROKODIL learners in form of direct communication or learning group membership as well as external connections, for example, in *Facebook* will be combined with indirect evidence such as bookmarking, viewing, or copying another user's resources. Based on preference information about tags, resources, and learning activities as well as structural information about the local social graph, different kinds of ties (like support, influence) and their strength will be identified. Those ties will be exploited to propagate relevant information and to recommend experts.

Bibliography

[1] Abdi, H. Kendall rank correlation. In *Encyclopedia of Measurement and Statistics*, N. J. Salkind, Ed. SAGE Publications, Inc., Thousand Oaks, California, USA, 2007, pp. 508–510.

[2] Adamic, L. A., and Adar, E. Friends and Neighbors on the Web. *Social Networks 25*, 3 (July 2003), 211–230.

[3] Adamic, L. A., Buyukkokten, O., and Adar, E. A Social Network Caught in the Web. *First Monday 8*, 6 (2003).

[4] Ames, M., and Naaman, M. Why We Tag: Motivations for Annotation in Mobile and Online Media. In *Proceedings of the 25th ACM SIGCHI International Conference on Human Factors in Computing Systems* (San Jose, California, USA, April 30-May 3 2007), CHI '07, ACM, pp. 971–980.

[5] Andrews, P., De Natale, F., Buschbeck, S., Jameson, A., Bischoff, K., Firan, C. S., Niederée, C., Mezaris, V., Nikolopoulos, S., Murdock, V., and Rae, A. GLOCAL: Event-based Retrieval of Networked Media. In *Proceedings of the 21st International Conference Companion on World Wide Web* (Lyon, France, April 16-20 2012), WWW '12 Companion, ACM, pp. 219–222.

[6] Anjorin, M., Rensing, C., Bischoff, K., Bogner, C., Lehmann, L., Reger, A. L., Faltin, N., Steinacker, A., Lüdemann, A., and Domínguez García, R. CROKODIL - A Platform for Collaborative Resource-Based Learning. In *Proceedings of the 6th European Conference of Technology Enhanced Learning* (Palermo, Italy, September 20-23 2011), EC-TEL '11, Springer-Verlag, pp. 29–42.

[7] Aral, S., Muchnik, L., and Sundararajan, A. Distinguishing Influence-based Contagion from Homophily-driven Diffusion in Dynamic Networks. *Proceedings of the National Academy of Sciences (PNAS) 106*, 51 (2009), 21544–21549.

[8] Au Yeung, C.-m., and Iwata, T. Strength of Social Influence in Trust Networks in Product Review Sites. In *Proceedings of the 4th International ACM Conference on Web Search and Data Mining* (Hong Kong, 2011), WSDM '11, ACM, pp. 495–504.

[9] Aurnhammer, M., Hanappe, P., and Steels, L. Integrating Collaborative Tagging and Emergent Semantics for Image Retrieval. In *Proceedings of the WWW Workshop on Collaborative Web Tagging* (Edinburgh, Scotland, UK, May 22 2006).

[10] Backstrom, L., Bakshy, E., Kleinberg, J. M., Lento, T. M., and Rosenn, I. Center of Attention: How Facebook Users Allocate Attention across Friends. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media* (Barcelona, Catalonia, Spain, July 17-21 2011), ICWSM '11, AAAI Press, pp. 34–41.

[11] Backstrom, L., Boldi, P., Rosa, M., Ugander, J., and Vigna, S. Four Degrees of Separation. In *Proceedings of the 4th International ACM Conference on Web Science* (Evanston, Illinois, USA, June 22-24 2012), WebSci '12, pp. 33–42.

[12] Backstrom, L., Huttenlocher, D., Kleinberg, J., and Lan, X. Group Formation in Large Social Networks: Membership, Growth, and Evolution. In *Proceedings of the 12th International ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (Philadelphia, Pennsylvania, USA, 2006), KDD '06, ACM, pp. 44–54.

[13] Backstrom, L., and Leskovec, J. Supervised Random Walks: Predicting and Recommending Links in Social Networks. In *Proceedings of the 4th International ACM Conference on Web Search and Data Mining* (Hong Kong, 2011), WSDM '11, ACM, pp. 635–644.

[14] Backstrom, L., Sun, E., and Marlow, C. Find Me If You Can: Improving Geographical Prediction with Social and Spatial Proximity. In *Proceedings of the 19th International Conference on World Wide Web* (Raleigh, North Carolina, USA, 2010), WWW '10, ACM, pp. 61–70.

[15] Bao, S., Xue, G., Wu, X., Yu, Y., Fei, B., and Su, Z. Optimizing Web Search Using Social Annotations. In *Proceedings of the 16th International Conference on World Wide Web* (Banff, Alberta, Canada, 2007), WWW '07, ACM, pp. 501–510.

[16] Bayma, N. K., and Ledbetter, A. Tunes that Bind? Predicting Friendship Strength in a Music-based Social Network. *Information, Communication & Society 12*, 3 (2009), 408–427.

[17] Bender, M., Crecelius, T., Kacimi, M., Michel, S., Neumann, T., Parreira, J. X., Schenkel, R., and Weikum, G. Exploiting social relations for query expansion and result ranking. In *Proceedings of the 24th International IEEE Conference on Data Engineering Workshop* (2008), ICDEW '08, IEEE Computer Society, pp. 501–506.

[18] Bertin-Mahieux, T., Eck, D., Maillet, F., and Lamere, P. Autotagger: A Model for Predicting Social Tags from Acoustic Features on Large Music Databases. *Journal of New Music Research 37*, 2 (2008), 115–135.

[19] Bischoff, K. Exploiting Social Ties for Search and Recommendation in Online Social Networks - Challenges and Chances. In *Proceedings des 22. Workshop über Grundlagen von Datenbanken des GI-Arbeitskreises Grundlagen von Informationssystemen* (Bad Helmstedt, Germany, May 25-28 2010), GvDB '10, CEUR Workshop Proceedings (CEUR-WS.org).

[20] Bischoff, K. We Love Rock 'n' Roll: Analyzing and Predicting Friendship Links in Last.fm. In *Proceedings of the ACM Web Science Conference* (Evanston, Illinois, USA, June 22-24 2012), WebSci '12, ACM, pp. 47–56. http://doi.acm.org/10.1145/2380718.2380725.

[21] Bischoff, K., Firan, C. S., Georgescu, M., Nejdl, W., and Paiu, R. Social Knowledge-Driven Music Hit Prediction. In *Proceedings of the 5th International Conference on Advanced Data Mining and Applications* (Beijing, China, August 17-19 2009), ADMA '09, Springer-Verlag, pp. 43–54.

[22] Bischoff, K., Firan, C. S., Kadar, C., Nejdl, W., and Paiu, R. Automatically Identifying Tag Types. In *Proceedings of the 5th International Conference on Advanced Data Mining and Applications* (Beijing, China, August 17-19 2009), ADMA '09, Springer-Verlag, pp. 31–42.

[23] Bischoff, K., Firan, C. S., Nejdl, W., and Paiu, R. Can All Tags be Used for Search? In *Proceedings of the 17th International ACM Conference on Information and Knowledge Management* (Napa Valley, California, USA, October 26-30 2008), CIKM '08, ACM, pp. 193–202.

[24] Bischoff, K., Firan, C. S., Nejdl, W., and Paiu, R. How Do You Feel About "Dancing Queen"?: Deriving Mood & Theme Annotations from User Tags. In *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital libraries* (Austin, Texas, USA, June 15-19 2009), JCDL '09, ACM, pp. 285–294.

[25] Bischoff, K., Firan, C. S., Nejdl, W., and Paiu, R. Bridging the Gap Between Tagging and Querying Vocabularies: Analyses and Applications for Enhancing Multimedia IR. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web 8*, 2-3 (July 2010), 97–109.

[26] Bischoff, K., Firan, C. S., and Paiu, R. Deriving Music Theme Annotations from User Tags. In *Proceedings of the 18th International Conference on World Wide Web* (Madrid, Spain, April 20-24 2009), WWW '09, ACM, pp. 1193–1194.

[27] Bischoff, K., Firan, C. S., Paiu, R., Nejdl, W., Laurier, C., and Sordo, M. Music Mood and Theme Classification - A Hybrid Approach. In *Proceedings of the 10th International Society for Music Information Retrieval Conference* (Kobe, Japan, October 26-30 2009), ISMIR '09, International Society for Music Information Retrieval, pp. 657–662.

[28] Bischoff, K., Herder, E., and Nejdl, W. Workplace Learning: How We Keep Track of Relevant Information. In *Proceedings of the 2nd European Conference on Technology Enhanced Learning* (Crete, Greece, September 17-20 2007), EC-TEL '07, Springer-Verlag, pp. 438–443.

[29] Bischoff, K., Mandl, T., Kölle, R., and Womser-Hacker, C. Geographische Bedingungen im Information Retrieval: Neue Ansätze in Systementwicklung und Evaluierung. In *Proceedings of the 10th International Symposium for Information Science* (Cologne, Germany, May 30-June 1 2007), ISI '07, UVK Verlagsgesellschaft mbH, Konstanz, Germany, pp. 15–26.

[30] Bischoff, K., Mandl, T., and Womser-Hacker, C. Blind Relevance Feedback and Named Entity Based Query Expansion for Geographic Retrieval at GeoCLEF 2006. In *Proceedings of the 7th Workshop of the Cross-Language Evaluation Forum* (Alicante, Spain, September 20-22 2006), CLEF '06, Springer-Verlag, pp. 946–953.

[31] Bischoff, K., Mandl, T., and Womser-Hacker, C. GeoCLEF 2006: Cross-linguales geographisches Information Retrieval. In *Proceedings of the German Society for Informatics (GI) Joint Workshop Event Lernen - Wissensentdeckung - Adaptivität, LWA 2006* (Hildesheim, Germany, October 9-11 2006), vol. 1/2006 of *Hildesheimer Informatik-Berichte*, University of Hildesheim, Institute of Computer Science, pp. 89–93.

[32] Blei, D. M., Ng, A. Y., and Jordan, M. I. Latent Dirichlet Allocation. *Journal of Machine Learning Research 3* (Mar. 2003), 993–1022.

[33] Boer, D., Fischer, R., Strack, M., Bond, M. H., Lo, E., and Lam, J. How Shared Preferences in Music Create Bonds Between People: Values as the Missing Link. *Personality and Social Psychology Bulletin 37*, 9 (2011), 1159–1171.

[34] Bonhard, P., Sasse, M. A., and Harries, C. "The Devil You Know Knows Best": How Online Recommendations Can Benefit from Social Networking. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...but not as we know it - Volume 1* (Lancaster, England, UK, September 3-7 2007), BCS-HCI '07, British Computer Society, pp. 77–86.

[35] Breiman, L. Bagging predictors. *Machine Learning 24*, 2 (Aug. 1996), 123–140.

[36] Broder, A. A Taxonomy of Web Search. *SIGIR Forum 36*, 2 (Sept. 2002), 3–10.

[37] Brooks, C. H., and Montanez, N. Improved Annotation of the Blogosphere via Autotagging and Hierarchical Clustering. In *Proceedings of the 15th International Conference on World Wide Web* (Edinburgh, Scotland, UK, May 23-26 2006), WWW '06, ACM, pp. 625–632.

[38] Brzozowski, M. J., Hogg, T., and Szabo, G. Friends and Foes: Ideological Social Networking. In *Proceedings of the 26th ACM SIGCHI International Conference on Human Factors in Computing Systems* (Florence, Italy, April 5-10 2008), CHI '08, ACM, pp. 817–820.

[39] Bundschus, M., Yu, S., Tresp, V., Rettinger, A., Dejori, M., and Kriegel, H.-P. Hierarchical Bayesian Models for Collaborative Tagging Systems. In *Proceedings of the 9th International IEEE Conference on Data Mining* (Miami, Florida ,USA, December 6-9 2009), ICDM '09, IEEE Computer Society, pp. 728–733.

[40] Burt, R. S. *Structural Holes: The Social Structure of Competition.* Harvard University Press, Cambridge, Massachusetts, USA, 1992.

[41] Burt, R. S. Structural Holes and Good Ideas. *American Journal of Sociology 110*, 2 (2004), 349–399.

[42] Byde, A., Wan, H., and Cayzer, S. Personalized Tag Recommendations via Tagging and Content-based Similarity Metrics. In *Proceedings of the 1st International Conference on Weblogs and Social Media* (Boulder, Colorado, USA, March 26-28 2007), ICWSM '07.

[43] Byron, L., Lento, T., Marlow, C., and Rosenn, I. Maintained Relationships on Facebook. online: `http://overstated.net/2009/03/09/maintained-relationships-on-facebook`, accessed: July 14, 2012, 2009.

[44] Cattuto, C., Loreto, V., and Pietronero, L. Semiotic Dynamics and Collaborative Tagging. *Proceedings of the National Academy of Sciences (PNAS) 104*, 5 (2007), 1461–1464.

[45] Cattuto, C., Schmitz, C., Baldassarri, A., Servedio, V. D. P., Loreto, V., Hotho, A., Grahl, M., and Stumme, G. Network Properties of Folksonomies. *Journal AI Communications - Network Analysis in Natural Sciences and Engineering 20*, 4 (Dec. 2007), 245–262.

[46] Chang, C.-C., and Lin, C.-J. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST) 2*, 3 (May 2011), 27:1–27:27.

[47] Chen, J., Geyer, W., Dugan, C., Muller, M., and Guy, I. Make New Friends, but Keep the Old: Recommending People on Social Networking Sites. In *Proceedings of the 27th ACM SIGCHI International Conference on Human Factors in Computing Systems* (Boston, Massachusetts, USA, April 4-9 2009), CHI '09, ACM, pp. 201–210.

[48] Chen, L., Wright, P., and Nejdl, W. Improving Music Genre Classification Using Collaborative Tagging Data. In *Proceedings of the 2nd International ACM Conference on Web Search and Data Mining* (Barcelona, Catalonia, Spain, February 9-12 2009), WSDM '09, ACM, pp. 84–93.

[49] Chen, W., and Fong, S. Social Network Collaborative Filtering Framework and Online Trust Factors: A Case Study on Facebook. In *Proceedings of the 5th International IEEE Conference on Digital Information Management* (Thunder Bay, Ontario, Canada, July 5-8 2010), ICDIM '10, IEEE Computer Society, pp. 266–273.

[50] Choy, S.-O., and Lui, A. K. Web Information Retrieval in Collaborative Tagging Systems. In *Proceedings of the International IEEE/WIC/ACM Conference on Web Intelligence* (December 18-22 2006), WI '06, IEEE Computer Society, pp. 352–355.

[51] Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement 20*, 1 (1960), 37–46.

[52] Condie, T., Kamvar, S. D., and Garcia-Molina, H. Adaptive Peer-to-Peer Topologies. In *Proceedings of the 4th International IEEE Conference on Peer-to-Peer Computing* (Zurich, Switzerland, August 25-27 2004), P2P '04, IEEE Computer Society, pp. 53–62.

[53] Corlette, D., and Shipman, F. M. Link Prediction Applied to an Open Large-scale Online Social Network. In *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia* (Toronto, Ontario, Canada, June 13-16 2010), HT '10, ACM, pp. 135–140.

[54] Cosley, D., Frankowski, D., Terveen, L., and Riedl, J. SuggestBot: Using Intelligent Task Routing to Help People Find Work in Wikipedia. In *Proceedings of the 12th International Conference on Intelligent User Interfaces* (Honolulu, Hawaii, USA, January 28-31 2007), IUI '07, ACM.

[55] Cosley, D., Ludford, P., and Terveen, L. Studying the Effect of Similarity in Online Task-focused Interactions. In *Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work* (Sanibel Island, Florida, USA, November 9-12 2003), GROUP '03, ACM, pp. 321–329.

[56] Cox, N. J. Speaking Stata: Correlation with confidence, or Fisher's z revisited. *The Stata Journal 8*, 3 (2008), 413–439.

[57] Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., and Suri, S. Feedback Effects Between Similarity and Social Influence in Online Communities. In *Proceedings of the 14th International ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (Las Vegas, Nevada, USA, August 24-27 2008), KDD '08, ACM, pp. 160–168.

[58] Crandall, D. J., Backstrom, L., Cosley, D., Suri, S., Huttenlocher, D., and Kleinberg, J. Inferring Social Ties from Geographic Coincidences. *Proceedings of the National Academy of Sciences (PNAS) 107*, 52 (2010), 22436–22441.

[59] Cristianini, N., and Shawe-Taylor, J. *An Introduction to Support Vector Machines and other kernel-based learning methods.* Cambridge University Press, 2000.

[60] danah boyd. Faceted Id/entity: Managing Representation in a Digital World. Master's thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, 2002. online: `http://www.danah.org/papers/Thesis.FacetedIdentity.pdf`.

[61] Davitz, J., Yu, J., Basu, S., Gutelius, D., and Harris, A. iLink: Search and Routing in Social Ntworks. In *Proceedings of the 13th International ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (San Jose, California, USA, August 12-15 2007), KDD '07, ACM, pp. 931–940.

[62] Demers, A., Greene, D., Hauser, C., Irish, W., Larson, J., Shenker, S., Sturgis, H., Swinehart, D., and Terry, D. Epidemic Algorithms for Replicated Database Maintenance. In *Proceedings of the 6th Annual ACM Symposium on Principles of Distributed Computing* (Vancouver, British Columbia, Canada, August 10-12 1987), PODC '87, ACM, pp. 1–12.

[63] Diehl, C. P., Namata, G., and Getoor, L. Relationship Identification for Social Network Discovery. In *Proceedings of the 22nd National Conference on Artificial Intelligence - Volume 1* (Vancouver, British Columbia, Canada, July 22-26 2007), AAAI '07, AAAI Press, pp. 546–552.

[64] Dubinko, M., Kumar, R., Magnani, J., Novak, J., Raghavan, P., and Tomkins, A. Visualizing Tags Over Time. In *Proceedings of the 15th International Conference on World Wide Web* (Edinburgh, Scotland, UK, May 23-26 2006), WWW '06, ACM, pp. 193–202.

[65] Dunbar, R. I. The Social Brain Hypothesis. *Evolutionary Anthropology 6*, 5 (1998), 178–190.

[66] Eagle, N., Pentland, A. S., and Lazer, D. Inferring Friendship Network Structure by Using Mobile Phone Data. *Proceedings of the National Academy of Sciences (PNAS) 106*, 36 (2009), 15274–15278.

[67] Easley, D., and Kleinberg, J. *Networks, Crowds, and Markets. Reasoning About a Highly Connected World.* Cambridge University Press, Cambridge, UK, 2010.

[68] Eck, D., Lamere, P., Bertin-Mahieux, T., and Green, S. Automatic Generation of Social Tags for Music Recommendation. In *Advances in Neural Information Processing Systems 20, Proceedings of the 21st Annual Conference on Neural Information Processing Systems* (Vancouver, British Columbia, Canada, December 3-6 2008), NIPS '07, MIT Press, pp. 385–392.

[69] Ehrlich, K., Lin, C.-Y., and Griffiths-Fisher, V. Searching for Experts in the Enterprise: Combining Text and Social Network Analysis. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work* (Sanibel Island, Florida, USA, November 4-7 2007), GROUP '07, ACM, pp. 117–126.

[70] Esuli, A., and Sebastiani, F. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. In *Proceedings of the 5th International Conference on Language Resources and Evaluation* (Genoa, Italy, May 24-26 2006), LREC '06, pp. 417–422.

[71] Evans, B. M., and Chi, E. H. Towards a Model of Understanding Social Search. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work* (San Diego, California, USA, November 8-12 2008), CSCW '08, ACM, pp. 485–494.

[72] Feng, Y., Zhuang, Y., and Pan, Y. Popular Music Retrieval by Detecting Mood. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval* (Toronto, Ontario, Canada, July 28-August 1 2003), SIGIR '03, ACM, pp. 375–376.

[73] Fiore, A. T., and Donath, J. S. Homophily in Online Dating: When Do You Like Someone Like Yourself? In *Extended Abstracts Proceedings of the 2005 Conference on Human Factors in Computing Systems* (Portland, Oregon, USA, April 2-7 2005), CHI EA '05, ACM, pp. 1371–1374.

[74] Firan, C. S., Nejdl, W., and Paiu, R. The Benefit of Using Tag-Based Profiles. In *Proceedings of the 2007 Latin American Web Conference* (Santiago de Chile, Chile, October 31-November 2 2007), LA-WEB '07, IEEE Computer Society, pp. 32–41.

[75] Fisher, R. A. *Statistical Methods for Research Workers.* Oliver & Boyd, Edinburgh, Scotland, UK, 1925.

[76] Fortunato, S. Community Detection in Graphs. *Physics Reports 486*, 3-5 (February 2010), 75–174.

[77] Frank, E., Wang, Y., Inglis, S., Holmes, G., and Witten, I. H. Using Model Trees for Classification. *Machine Learning 32* (1998), 63–76.

[78] Frank J. Massey, J. The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association 46*, 253 (1951), 68–78.

[79] Franke, T. M., Ho, T., and Christie, C. A. The Chi-Square Test. *American Journal of Evaluation 33*, 3 (2012), 448–458.

[80] Friedkin, N. E. A Test of Structural Features of Granovetter's Strength of Weak Ties Theory. *Social Networks 2*, 4 (1980), 411–422.

[81] Furnas, G. W., Landauer, T. K., Gomez, L. M., and Dumais, S. T. The Vocabulary Problem in Human-System Communication. *Communications of the ACM 30*, 11 (November 1987), 964–971.

[82] Gardikiotis, A., and Baltzis, A. 'Rock Music for Myself and Justice to the World!': Musical Identity, Values, and Music Preferences. *Psychology of Music 40*, 2 (2012), 143–163.

[83] Gey, F., Larson, R., Sanderson, M., Bischoff, K., Mandl, T., Womser-Hacker, C., Santos, D., Rocha, P., Di Nunzio, G. M., and Ferro, N. Challenges to Evaluation of Multilingual Geographic Information Retrieval in GeoCLEF. In *Proceedings of the 1st International Workshop on Evaluating Information Access* (Tokyo, Japan, May 15 2007), EVIA '07, `http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings6/EVIA`, pp. 74–77.

[84] Gey, F. C., Larson, R. R., Sanderson, M., Bischoff, K., Mandl, T., Womser-Hacker, C., Santos, D., Rocha, P., Di Nunzio, G. M., and Ferro, N. GeoCLEF 2006: The CLEF 2006 Cross-Language Geographic Information Retrieval Track Overview. In *Proceedings of the 7th Workshop of the Cross-Language Evaluation Forum* (Alicante, Spain, September 20-22 2006), CLEF '06, Springer-Verlag, pp. 852–876.

[85] Gibbons, J. D., and Chakraborti, S. *Nonparametric Statistical Inference*, 5 ed. Chapman and Hall/CRC Press, Taylor & Francis Group, Boca Raton, Florida, USA, 2010.

[86] Gilbert, E. Predicting Tie Strength in a New Medium. In *Proceedings of the 2012 ACM Conference on Computer Supported Cooperative Work* (Seattle, Washington, USA, February 11-15 2012), CSCW '12, ACM, pp. 1047–1056.

[87] Gilbert, E., and Karahalios, K. Predicting Tie Strength with Social Media. In *Proceedings of the 27th ACM SIGCHI International Conference on Human Factors in Computing Systems* (Boston, Massachusetts, USA, April 4-9 2009), CHI '09, ACM, pp. 211–220.

[88] Giles, J. Internet encyclopaedias go head to head. *Nature 438*, 900-901 (2005).

[89] Girvan, M., and Newman, M. E. J. Community Structure in Social and Biological Networks. *Proceedings of the National Academy of Sciences (PNAS) 99*, 12 (2002), 7821–7826.

[90] Golder, S. A., and Huberman, B. A. Usage Patterns of Collaborative Tagging Systems. *Journal of Information Science 32*, 2 (2006), 198–208.

[91] Golder, S. A., and Yardi, S. Structural Predictors of Tie Formation in Twitter: Transitivity and Mutuality. In *Proceedings of the 2nd International IEEE Conference on Social Computing* (Minneapolis, Minnesota, USA, August 20-22 2010), SocialCom '10, IEEE Computer Society, pp. 88–95.

[92] Granovetter, M. S. The Strength of Weak Ties. *American Journal of Sociology 78*, 6 (1973), 1360–1380.

[93] Granovetter, M. S. The Strength of Weak Ties: a Network Theory Revisited. *Sociological Theory 1* (1983), 201–233.

[94] Groh, G., and Ehmig, C. Recommendations in Taste Related Domains: Collaborative Filtering vs. Social Filtering. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work* (Sanibel Island, Florida, USA, November 4-7 2007), GROUP '07, ACM, pp. 127–136.

[95] Gruhl, D., Guha, R., Liben-Nowell, D., and Tomkins, A. Information Diffusion Through Blogspace. In *Proceedings of the 13th International Conference on World Wide Web* (New York, New York, USA, May 17-22 2004), WWW '04, ACM, pp. 491–501.

[96] Gupte, M., and Eliassi-Rad, T. Measuring Tie Strength in Implicit Social Networks. In *Proceedings of the 4th International ACM Conference on Web Science* (Evanston, Illinois, USA, June 22-24 2012), WebSci '12, pp. 109–118.

[97] Guy, I., Jacovi, M., Perer, A., Ronen, I., and Uziel, E. Same Places, Same Things, Same People?: Mining User Similarity on Social Media. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (Savannah, Georgia, USA, February 6-10 2010), CSCW '10, ACM, pp. 41–50.

[98] Guy, I., Jacovi, M., Shahar, E., Meshulam, N., Soroka, V., and Farrell, S. Harvesting with SONAR: The Value of Aggregating Social Network Information. In *Proceedings of the 26th ACM SIGCHI International Conference on Human Factors in Computing Systems* (Florence, Italy, April 5-10 2008), CHI '08, ACM, pp. 1017–1026.

[99] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. The WEKA Data Mining Software: an Update. *ACM SIGKDD Explorations Newsletter 11*, 1 (Nov. 2009), 10–18.

[100] Hall, M. A., and Smith, L. A. Feature Subset Selection: a Correlation based Filter Approach. In *Proceedings of the 4th International Conference on Neural Information Processing and Intelligent Information Systems* (Dunedin, New Zealand, 1997), pp. 855–858.

[101] Halpin, H., Robu, V., and Shepherd, H. The Complex Dynamics of Collaborative Tagging. In *Proceedings of the 16th International Conference on World Wide Web* (Banff, Alberta, Canada, May 8-12 2007), WWW '07, ACM, pp. 211–220.

[102] Hannafin, M. J., and Hill, J. R. Resource-based learning. In *Handbook of Research on Educational Communications and Technology*, J. M. Spector, M. D. Merrill, J. van Merrienboer, and M. P. Driscoll, Eds. Erlbaum, Mahwah, New Jersey, USA, 2007, pp. 525–536.

[103] Heckner, M., Heilemann, M., and Wolff, C. Personal Information Management vs. Resource Sharing: Towards a Model of Information Behavior in Social Tagging Systems. In *Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media* (San Jose, California, USA, May 17-20 2009), ICWSM '09, AAAI Press, pp. 42–49.

[104] Heckner, M., Mühlbacher, S., and Wolff, C. Tagging Tagging. Analysing User Keywords in Scientific Bibliography Management Systems. *Journal of Digital Information 9*, 2 (2008).

[105] Heckner, M., Neubauer, T., and Wolff, C. Tree, funny, to_read, google: What are Tags Supposed to Achieve? A Comparative Analysis of User Keywords for Different Digital Resource Types. In *Proceedings of the 2008 ACM Workshop on Search in Social Media* (Napa Valley, California, USA, October 30 2008), SSM '08, ACM, pp. 3–10.

[106] Heymann, P., Koutrika, G., and Garcia-Molina, H. Can Social Bookmarking Improve Web Search? In *Proceedings of the 1st International ACM Conference on Web Search and Data Mining* (Palo Alto, California, USA, February 11-12 2008), WSDM '08, ACM, pp. 195–206.

[107] Heymann, P., Ramage, D., and Garcia-Molina, H. Social Tag Prediction. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Singapore, Singapore, July 20-24 2008), SIGIR '08, ACM, pp. 531–538.

[108] Hollander, M., and Wolfe, D. A. *Nonparametric Statistical Methods*, 2 ed. Wiley-Interscience, 1999.

[109] Hopcroft, J., Lou, T., and Tang, J. Who Will Follow You Back?: Reciprocal Relationship Prediction. In *Proceedings of the 20th International ACM Conference on Information and Knowledge Management* (Glasgow, Scotland, UK, October 24-28 2011), CIKM '11, ACM, pp. 1137–1146.

[110] Horowitz, D., and Kamvar, S. D. The Anatomy of a Large-scale Social Search Engine. In *Proceedings of the 19th International Conference on World Wide Web* (Raleigh, North Carolina, USA, April 26-30 2010), WWW '10, ACM, pp. 431–440.

[111] Hotho, A., Jäschke, R., Schmitz, C., and Stumme, G. Information Retrieval in Folksonomies: Search and Ranking. In *Proceedings of the 3rd European Conference on the Semantic Web: Research and Applications* (Budva, Montenegro, June 11-14 2006), ESWC '06, Springer-Verlag, pp. 411–426.

[112] Hu, X., and Downie, J. S. Exploring Mood Metadata: Relationships with Genre, Artist and Usage Metadata. In *Proceedings of the 8th International Conference on Music Information Retrieval* (Vienna, Austria, September 23-27 2007), ISMIR '07, Austrian Computer Society, pp. 67–72.

[113] Huberman, B. A., Romero, D. M., and Wu, F. Social Networks that Matter: Twitter Under the Microscope. *First Monday 14*, 1 (January 2009).

[114] Indra Mohan Chakravarti, R. G. Laha, J. R. *Handbook of Methods of Applied Statistics, Volume I: Techniques of Computation, Descriptive Methods, and Statistical Inference.* John Wiley and Sons, 1967.

[115] Iofciu, T., Fankhauser, P., Abel, F., and Bischoff, K. Identifying Users Across Social Tagging Systems. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media* (Barcelona, Catalonia, Spain, July 17-21 2011), ICWSM '11, AAAI Press, pp. 522–525.

[116] Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., and Stumme, G. Tag Recommendations in Social Bookmarking Systems. *AI Communications 21*, 4 (Dec. 2008), 231–247.

[117] Kahanda, I., and Neville, J. Using Transactional Information to Predict Link Strength in Online Social Networks. In *Proceedings of the 3rd AAAI International Conference on Weblogs and Social Media* (San Jose, California, USA, May 17-20 2009), ICWSM '09, AAAI Press, pp. 74–81.

[118] Kamvar, S. D., Schlosser, M. T., and Garcia-Molina, H. The Eigentrust Algorithm for Reputation Management in P2P networks. In *Proceedings*

of the 12th International Conference on World Wide Web (Budapest, Hungary, May 20-24 2003), WWW '03, ACM, pp. 640–651.

[119] Karweg, B., Huetter, C., and Böhm, K. Evolving Social Search Based on Bookmarks and Status Messages from Social Networks. In *Proceedings of the 20th International ACM Conference on Information and Knowledge Management* (Glasgow, Scotland, UK, October 24-28 2011), CIKM '11, ACM, pp. 1825–1834.

[120] Katz, L. A New Status Index Derived from Sociometric Analysis. *Psychometrika 18*, 1 (March 1953), 39–43.

[121] Kautz, H., Selman, B., and Shah, M. Referral Web: Combining Social Networks and Collaborative Filtering. *Communications of the ACM 40*, 3 (March 1997), 63–65.

[122] Kazienko, P., and Musial, K. Recommendation Framework for Online Social Networks. In *Advances in Web Intelligence and Data Mining*, M. Last, P. S. Szczepaniak, Z. Volkovich, and A. Kandel, Eds., vol. 23 of *Studies in Computational Intelligence*. Springer-Verlag, Berlin / Heidelberg, Germany, 2006, pp. 111–120.

[123] Kendall, M. G. *Rank Correlation Methods*, 4 ed. Griffin, London, England, UK, 1970.

[124] Kipp, M. E. I., and Campbell, D. G. Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices. *Proceedings of the American Society for Information Science and Technology 43*, 1 (2006), 1–18.

[125] Kittur, A., Suh, B., Pendleton, B. A., and Chi, E. H. He says, she says: conflict and coordination in Wikipedia. In *Proceedings of the 25th ACM SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA, 2007), CHI '07, ACM, pp. 453–462.

[126] Kleinberg, J. M. Authoritative Sources in a Hyperlinked Environment. *J. ACM 46*, 5 (1999), 604–632.

[127] Krestel, R., and Fankhauser, P. Personalized Topic-based Tag Recommendation. *Neurocomputing 76*, 1 (Jan. 2012), 61–70.

[128] Krestel, R., Fankhauser, P., and Nejdl, W. Latent Dirichlet Allocation for Tag Recommendation. In *Proceedings of the 3rd International ACM Conference on Recommender Systems* (New York, New York, USA, September 9-13 2009), RecSys '09, ACM, pp. 61–68.

[129] La Fond, T., and Neville, J. Randomization Tests for Distinguishing Social Influence and Homophily Effects. In *Proceedings of the 19th International Conference on World Wide Web* (Raleigh, North Carolina, USA, April 26-30 2010), WWW '10, ACM, pp. 601–610.

[130] Laurier, C., and Herrera, P. Audio Music Mood Classification Using Support Vector Machine. In *Proceedings of the 8th International Conference on Music Information Retrieval* (Vienna, Austria, September 23-27 2007), ISMIR '07, Austrian Computer Society.

[131] Laurier, C., and Herrera, P. Automatic Detection of Emotion in Music: Interaction with Emotionally Sensitive Machines. In *Handbook of Research on Synthetic Emotions and Sociable Robotics: New Applications in Affective Computing and Artificial Intelligence*, J. Vallverdú and D. Casacuberta, Eds. IGI Global, Hershey, Pennsylvania, USA, 2009, ch. 2, pp. 9–32.

[132] Lee, D. H., and Brusilovsky, P. Social Networks and Interest Similarity: The Case of CiteULike. In *Proceedings of the 21st ACM Conference on Hypertext and Hypermedia* (Toronto, Ontario, Canada, June 13-16 2010), HT '10, ACM, pp. 151–156.

[133] Lee, J. H., and Downie, J. S. Survey Of Music Information Needs, Uses, And Seeking Behaviours: Preliminary Findings. In *Proceedings of the 5th International Conference on Music Information Retrieval* (Barcelona, Catalonia, Spain, October 10-15 2004), ISMIR '04.

[134] Leskovec, J., and Horvitz, E. Planetary-scale Views on a Large Instant-messaging Network. In *Proceedings of the 17th International Conference on World Wide Web* (Beijing, China, April 21-25 2008), WWW '08, ACM, pp. 915–924.

[135] Leskovec, J., Huttenlocher, D., and Kleinberg, J. Predicting Positive and Negative Links in Online Social Networks. In *Proceedings of the 19th International Conference on World Wide Web* (Raleigh, North Carolina, USA, April 26-30 2010), WWW '10, ACM, pp. 641–650.

[136] Levene, H. Robust tests for equality of variances. In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, I. Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, and H. B. Mann, Eds. Stanford University Press, 1960, pp. 278–292.

[137] Levy, M., and Sandler, M. A Semantic Space for Music Derived from Social Tags. In *Proceedings of the 8th International Conference on Music Information Retrieval* (Vienna, Austria, September 23-27 2007), ISMIR '07, Austrian Computer Society, pp. 411–416.

[138] Levy, M., and Sandler, M. Learning Latent Semantic Models for Music from Social Tags. *Journal of New Music Research 37*, 2 (2008), 137–150.

[139] Levy, M., and Sandler, M. Music Information Retrieval Using Social Tags and Audio. *IEEE Transactions on Multimedia - Special section on communities and media computing 11*, 3 (April 2009), 383–395.

[140] Lewis, D., and Burke, C. J. The Use and Misuse of the Chi-square Test. *Psychological Bulletin 46*, 6 (1949), 433–489.

[141] Li, X., Guo, L., and Zhao, Y. E. Tag-based Social Interest Discovery. In *Proceedings of the 17th International Conference on World Wide Web* (Beijing, China, April 21-25 2008), WWW '08, ACM, pp. 675–684.

[142] Liben-Nowell, D., and Kleinberg, J. The Link Prediction Problem for Social Networks. In *Proceedings of the 12th International ACM Conference on Information and Knowledge Management* (New Orleans, Louisiana, USA, November 2-8 2003), CIKM '03, ACM, pp. 556–559.

[143] Lichtenwalter, R. N., Lussier, J. T., and Chawla, N. V. New Perspectives and Methods in Link Prediction. In *Proceedings of the 16th International ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (Washington, District of Columbia, USA, July 25-28 2010), KDD '10, ACM, pp. 243–252.

[144] Lilliefors, H. W. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *Journal of the American Statistical Association 62*, 318 (1967), 399–402.

[145] Lin, N., Dayton, P. W., and Greenwald, P. Analyzing the Instrumental Use of Relations in the Context of Social Structure. *Sociological Methods & Research 7*, 2 (November 1978), 149–166.

[146] Lin, N., Ensel, W. M., and Vaughn, J. C. Social Resources and Strength of Ties: Structural Factors in Occupational Status Attainment. *American Sociological Review 46*, 4 (August 1981), 393–405.

[147] Liu, D., Lu, L., and Zhang, H. Automatic Mood Detection from Acoustic Music Data. In *Proceedings of the 4th International Conference on Music Information Retrieval* (Baltimore, Maryland, USA, 26-30 October 2003), ISMIR '03.

[148] Liu, H. Social Network Profiles as Taste Performances. *Journal of Computer-Mediated Communication 13*, 1 (October 2007), 252–275.

[149] Ma, H., Yang, H., Lyu, M. R., and King, I. SoRec: Social Recommendation Using Probabilistic Matrix Factorization. In *Proceedings of the 17th International ACM Conference on Information and Knowledge Management* (Napa Valley, California, USA, October 26-30 2008), CIKM '08, ACM, pp. 931–940.

[150] Macgregor, G., and McCulloch, E. Collaborative Tagging as a Knowledge Organisation and Resource Discovery Tool. *Library Review 55*, 5 (2006), 291–300.

[151] Macskassy, S. A., and Michelson, M. Why do People Retweet? Anti-Homophily Wins the Day! In *Proceedings of the 5th AAAI International Conference on Weblogs and Social Media* (Barcelona, Catalonia, Spain, July 17-21 2011), ICWSM '11, AAAI Press, pp. 209–216.

[152] Manning, C. D., Raghavan, P., and Schütze, H. *Introduction to Information Retrieval.* Cambridge University Press, 2008.

[153] Marchetti, A., Tesconi, M., Ronzano, F., Rosella, M., and Minutoli, S. SemKey: A Semantic Collaborative Tagging System. In *Proceedings of the WWW Workshop on Tagging and Metadata for Social Information Organization* (Banff, Alberta, Canada, May 8 2007).

[154] Marinho, L. B., Nanopoulos, A., Schmidt-Thieme, L., Jäschke, R., Hotho, A., Stumme, G., and Symeonidis, P. Social Tagging Recommender Systems. In *Recommender Systems Handbook*, F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, Eds. Springer-Verlag, Berlin / Heidelberg, Germany, 2011, pp. 615–644.

[155] Marlow, C., Naaman, M., Boyd, D., and Davis, M. HT06, tagging paper, taxonomy, Flickr, academic article, to read. In *Proceedings of the 17th Conference on Hypertext and Hypermedia* (Odense, Denmark, August 23-25 2006), HYPERTEXT '06, ACM, pp. 31–40.

[156] Marsden, P. V., and Campbell, K. E. Measuring Tie Strength. *Social Forces 63*, 2 (December 1984), 482–501.

[157] Mathes, A. Folksonomies - Cooperative Classification and Communication Through Shared Metadata. University of Illinois Urbana-Champaign Technical Report. online: `http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html`, accessed: August 20, 2012, 2004.

[158] McAfee, A. How to Hit the Enterprise 2.0 Bullseye. online: `http://andrewmcafee.org/2007/11/how_to_hit_the_enterprise_20_bullseye`, accessed: August 20, 2012, 2007.

[159] Mccallum, A., and Nigam, K. A Comparison of Event Models for Naive Bayes Text Classification. In *Proceedings of the AAAI 1998 Workshop on Learning for Text Categorization* (1998), pp. 41–48.

[160] McPherson, M., Smith-Lovin, L., and Cook, J. M. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology 27* (August 2001), 415–444.

[161] Meeder, B., Karrer, B., Sayedi, A., Ravi, R., Borgs, C., and Chayes, J. We Know Who You Followed Last Summer: Inferring Social Link Creation Times in Twitter. In *Proceedings of the 20th International Conference on World Wide Web* (Hyderabad, India, March 28-April 1 2011), WWW '11, ACM, pp. 517–526.

[162] Mishne, G. AutoTag: A Collaborative Approach to Automated Tag Assignment for Weblog Posts. In *Proceedings of the 15th International Conference on World Wide Web* (Edinburgh, Scotland, UK, May 23-26 2006), WWW '06, ACM, pp. 953–954.

[163] Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., and Bhattacharjee, B. Measurement and Analysis of Online Social Networks. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement* (San Diego, California, USA, October 24-26 2007), IMC '07, ACM, pp. 29–42.

[164] Mislove, A., Viswanath, B., Gummadi, K. P., and Druschel, P. You Are Who You Know: Inferring User Profiles in Online Social Networks. In *Proceedings of the 3rd International ACM Conference on Web Search and Data Mining* (New York, New York, USA, February 3-6 2010), WSDM '10, ACM, pp. 251–260.

[165] Morris, M. R., Teevan, J., and Panovich, K. What Do People ask Their Social Networks, and Why?: A Survey Study of Status Message q&a Behavior. In *Proceedings of the 28th ACM SIGCHI International Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA, April 10-15 2010), CHI '10, ACM, pp. 1739–1748.

[166] Muralidharan, A., Gyongyi, Z., and Chi, E. Social Annotations in Web Search. In *Proceedings of the 30th ACM SIGCHI International Conference on Human Factors in Computing Systems* (Austin, Texas, USA, 2012), CHI '12, ACM, pp. 1085–1094.

[167] Myers, E. W. An O(ND) Difference Algorithm and Its Variations. *Algorithmica 1* (1986), 251–266.

[168] Naaman, M., and Nair, R. ZoneTag's Collaborative Tag Suggestions: What is This Person Doing in My Phone? *IEEE MultiMedia 15*, 3 (July 2008), 34–40.

[169] Nakamoto, R., Nakajima, S., Miyazaki, J., and Uemura, S. Tag-Based Contextual Collaborative Filtering. *IAENG International Journal of Computer Science 34*, 2 (2008), 214–219.

[170] Nakamoto, R. Y., Nakajima, S., Miyazaki, J., Uemura, S., Kato, H., and Inagaki, Y. Reasonable Tag-based Collaborative Filtering for Social Tagging Systems. In *Proceedings of the 2nd ACM Workshop on Information Credibility on the Web* (Napa Valley, California, USA, October 30 2008), WICOW '08, ACM, pp. 11–18.

[171] Newman, M. E. J., and Girvan, M. Finding and Evaluating Community Structure in Networks. *Physical Review E 69*, 2 (February 2004).

[172] Nov, O. What motivates Wikipedians? *Communications of the ACM 50*, 11 (Nov. 2007), 60–64.

[173] Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D. M. J., Kaski, K., Kertész, J., and Barabási, A.-L. Structure and Tie Strengths in Mobile Communication Networks. *Proceedings of the National Academy of Sciences (PNAS) 104*, 18 (2008), 7332–7336.

[174] Overell, S., Sigurbjörnsson, B., and van Zwol, R. Classifying Tags Using Open Content Resources. In *Proceedings of the 2nd International ACM Conference on Web Search and Data Mining* (Barcelona, Catalonia, Spain, February 9-12 2009), WSDM '09, ACM, pp. 64–73.

[175] Page, L., Brin, S., Motwani, R., and Winograd, T. The PageRank Citation Ranking: Bringing Order to the Web. Technical Report 1999-66, Stanford InfoLab, November 1999. Previous number = SIDL-WP-1999-0120.

[176] Pass, G., Chowdhury, A., and Torgeson, C. A Picture of Search. In *Proceedings of the 1st international Conference on Scalable Information Systems* (Hong Kong, May 29-June 1 2006), InfoScale '06, ACM.

[177] Pazzani, M. J., and Billsus, D. Content-based Recommendation Systems. In *The Adaptive Web*, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds., LNCS 4321. Springer-Verlag, 2007, pp. 325–341.

[178] Quinlan, R. *C4.5: Programs for Machine Learning.* Morgan Kaufmann Publishers, San Mateo, California, USA, 1993.

[179] Rattenbury, T., Good, N., and Naaman, M. Towards Automatic Extraction of Event and Place Semantics from Flickr Tags. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Amsterdam, The Netherlands, July 23-27 2007), SIGIR '07, ACM, pp. 103–110.

[180] Rentfrow, P. J., and Gosling, S. D. The Do Re Mi's of Everyday Life: The Structure and Personality Correlates of Music Preferences. *Journal of Personality and Social Psychology 84*, 6 (June 2003), 1236–1256.

[181] Rentfrow, P. J., McDonald, J. A., and Oldmeadow, J. A. You Are What You Listen To: Young People's Stereotypes about Music Fans. *Group Processes & Intergroup Relations 12*, 3 (May 2009), 329–344.

[182] Rivera, M. T., Soderstrom, S. B., and Uzzi, B. Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms. *Annual Review of Sociology 36* (August 2010), 91–115.

[183] Salton, G., and McGill, M. J. *Introduction to Modern Information Retrieval.* McGraw-Hill, Inc., New York, New York, USA, 1986.

[184] Schafer, J. B., Frankowski, D., Herlocker, J., and Sen, S. Collaborative Filtering Recommender Systems. In *The Adaptive Web*, P. Brusilovsky, A. Kobsa, and W. Nejdl, Eds., LNCS 4321. Springer-Verlag, 2007, pp. 291–324.

[185] Schenkel, R., Crecelius, T., Kacimi, M., Michel, S., Neumann, T., Parreira, J. X., and Weikum, G. Efficient Top-k Querying over Social-tagging Networks. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Singapore, Singapore, July 20-24 2008), SIGIR '08, ACM, pp. 523–530.

[186] Schenkel, R., Crecelius, T., Kacimi, M., Neumann, T., Parreira, J. X., Spaniol, M., and Weikum, G. Social Wisdom for Search and Recommendation. *IEEE Data Engineering Bulletin 31*, 2 (June 2008), 40–49.

[187] Schifanella, R., Barrat, A., Cattuto, C., Markines, B., and Menczer, F. Folks in Folksonomies: Social Link Prediction from Shared Metadata. In *Proceedings of the 3rd International ACM Conference on Web Search and Data Mining* (New York, New York, USA, February 3-6 2010), WSDM '10, pp. 271–280.

[188] Schmitz, P. Inducing Ontology from Flickr Tags. In *Proceedings of the WWW Workshop on Collaborative Tagging* (Edinburgh, Scotland, UK, May 22 2006).

[189] Scholl, P., Mann, D., Rensing, C., and Steinmetz, R. Support of Acquisition and Organization of Kowledge Artifacts in Informal Learning Contexts. In *Proceedings of the EDEN Annual Conference: New Learning 2.0?* (Naples, Italy, June 13-16 2007), p. 16.

[190] Sen, S., Harper, F. M., LaPitz, A., and Riedl, J. The Quest for Quality Tags. In *Proceedings of the 2007 International ACM Conference on Supporting Group Work* (Sanibel Island, Florida, USA, November 4-7 2007), GROUP '07, ACM, pp. 361–370.

[191] Sen, S., Lam, S. K., Rashid, A. M., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F. M., and Riedl, J. tagging, communities, vocabulary, evolution. In *Proceedings of the 2006 ACM Conference on Computer Supported Cooperative Work* (Banff, Alberta, Canada, November 4-8 2006), CSCW '06, ACM, pp. 181–190.

[192] Sen, S., Vig, J., and Riedl, J. Learning to Recognize Valuable Tags. In *Proceedings of the 14th International Conference on Intelligent User Interfaces* (Sanibel Island, Florida, USA, February 8-11 2009), IUI '09, ACM, pp. 87–96.

[193] Sen, S., Vig, J., and Riedl, J. Tagommenders: Connecting Users to Items Through Tags. In *Proceedings of the 18th International Conference on World Wide Web* (Madrid, Spain, April 20-24 2009), WWW '09, ACM, pp. 671–680.

[194] Servia Rodríguez, S., Redondo, R. P. D., Fernández Vilas, A., and Pazos Arias, J. J. Using Facebook Activity to Infer Social Ties. In *Proceedings of the 2nd International Conference on Cloud Computing and Services Science* (Porto, Portugal, April 18-21 2012), CLOSER '12, SciTePress, pp. 325–333.

[195] Seth, A., and Zhang, J. A Social Network Based Approach to Personalized Recommendation of Participatory Media Content. In *Proceedings of the 2nd International AAAI Conference on Weblogs and Social Media* (Seattle, Washington, USA, March 30-April 2 2008), ICWSM '08, AAAI Press, pp. 109–117.

[196] Shannon, C. E. A Mathematical Theory of Communication. *The Bell System Technical Journal 27* (1948), 379–423 & 623–656.

[197] Shaver, P., Schwartz, J., Kirson, D., and O'Connor, C. Emotion Knowledge: Further Exploration of a Prototype Approach. *Journal of Personality and Social Psychology 52*, 6 (June 1987), 1061–1086.

[198] Shi, X., Adamic, L. A., and Strauss, M. J. Networks of Strong Ties. *Physica A: Statistical Mechanics and its Applications 378*, 1 (May 2007), 33–47.

[199] Sigurbjörnsson, B., and van Zwol, R. Flickr Tag Recommendation Based on Collective Knowledge. In *Proceedings of the 17th International Conference*

on World Wide Web (Beijing, China, April 21-25 2008), WWW '08, ACM, pp. 327–336.

[200] Sinha, R. R., and Swearingen, K. Comparing Recommendations Made by Online Systems and Friends. In *Proceedings of the 2nd DELOS Network of Excellence Workshop on Personalisation and Recommender Systems in Digital Libraries* (Dublin, Ireland, June 18-20 2001).

[201] Sinnott, R. W. Virtues of the Haversine. *Sky and Telescope 68*, 2 (1984), 159.

[202] Skowronek, J., McKinney, M. F., and van de Par, S. A Demonstrator for Automatic Music Mood Estimation. In *Proceedings of the 8th International Conference on Music Information Retrieval* (Vienna, Austria, September 23-27 2007), ISMIR '07, Austrian Computer Society, pp. 345–346.

[203] Song, X., Chi, Y., Hino, K., and Tseng, B. L. Information Flow Modeling Based on Diffusion Rate for Prediction and Ranking. In *Proceedings of the 16th International Conference on World Wide Web* (Banff, Alberta, Canada, May 8-12 2007), WWW '07, ACM, pp. 191–200.

[204] Sood, S., Owsley, S., Hammond, K., and Birnbaum, L. TagAssist: Automatic Tag Suggestion for Blog Posts. In *Proceedings of the 1st International Conference on Weblogs and Social Media* (Boulder, Colorado, USA, March 26-28 2007), ICWSM '07.

[205] Symeonidis, P., Nanopoulos, A., and Manolopoulos, Y. Tag Recommendations Based on Tensor Dimensionality Reduction. In *Proceedings of the 2nd International ACM Conference on Recommender Systems* (Lausanne, Switzerland, October 23-25 2008), RecSys '08, ACM, pp. 43–50.

[206] Symeonidis, P., Ruxanda, M. M., Nanopoulos, A., and Manolopoulos, Y. Ternary Semantic Analysis of Social Tags for Personalized Music Recommendation. In *Proceedings of the 9th International Conference on Music Information Retrieval* (Philadelphia, Pennsylvania, USA, September 14-18 2008), ISMIR '08, pp. 219–224.

[207] Tang, J., Lou, T., and Kleinberg, J. Inferring Social Ties Across Heterogenous Networks. In *Proceedings of the 5th International ACM Conference on Web Search and Data Mining* (Seattle, Washington, USA, February 8-12 2012), WSDM '12, ACM, pp. 743–752.

[208] Tang, W., Zhuang, H., and Tang, J. Learning to Infer Social Ties in Large Networks. In *Proceedings of the 2011 European Conference on Machine Learning and Knowledge Discovery in Databases - Volume Part III* (Athens, Greece, September 5-9 2011), ECML PKDD '11, Springer-Verlag, pp. 381–397.

[209] Thayer, R. E. *The Biopsychology of Mood and Arousal.* Oxford University Press, New York, New York, USA, 1990.

[210] Thelwall, M. Homophily in MySpace. *Journal of the American Society for Information Science and Technology 60*, 2 (February 2009), 219–231.

[211] Tran, N. K., Zerr, S., Bischoff, K., Niederee, C., and Krestel, R. Topic cropping: Leveraging latent topics for the analysis of small corpora. In *Proceedings of the 17th International Conference on Theory and Practice of Digital Libraries* (Valletta, Malta, September 22-26 2013), TPDL 2013, Springer-Verlag.

[212] Travers, J., and Milgram, S. An Experimental Study of the Small World Problem. *Sociometry 32*, 4 (December 1969), 425–443.

[213] Tso-Sutter, K. H. L., Marinho, L. B., and Schmidt-Thieme, L. Tag-aware Recommender Systems by Fusion of Collaborative Filtering Algorithms. In *Proceedings of the 23rd Annual ACM Symposium on Applied Computing* (Fortaleza, Ceara, Brazil, March 16-20 2008), SAC '08, ACM, pp. 1995–1999.

[214] Turnbull, D., Barrington, L., and Lanckriet, G. R. G. Modeling Music and Words Using a Multi-class Naïve Bayes Approach. In *Proceedings of the 7th International Conference on Music Information Retrieval* (Victoria, British Columbia, Canada, October 8-12 2006), ISMIR '06, pp. 254–259.

[215] Turnbull, D., Barrington, L., and Lanckriet, G. R. G. Five Approaches to Collecting Tags for Music. In *Proceedings of the 9th International Conference on Music Information Retrieval* (Philadelphia, Pennsylvania, USA, September 14-18 2008), ISMIR '08, pp. 225–230.

[216] Verbrugge, L. M. The Structure of Adult Friendship Choices. *Social Forces 56*, 2 (December 1977), 576–597.

[217] Viégas, F. B., Wattenberg, M., and Dave, K. Studying cooperation and conflict between authors with history flow visualizations. In *Proceedings of the 22nd ACM SIGCHI Conference on Human Factors in Computing Systems* (Vienna, Austria, 2004), CHI '04, ACM, pp. 575–582.

[218] Viegas, F. B., Wattenberg, M., Kriss, J., and van Ham, F. Talk Before You Type: Coordination in Wikipedia. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences* (2007), HICSS '07, pp. 78–87.

[219] von Ahn, L., and Dabbish, L. Designing Games with a Purpose. *Communications of the ACM 51*, 8 (Aug. 2008), 58–67.

[220] Voss, J. Tagging, Folksonomy & Co - Renaissance of Manual Indexing? *CoRR abs/cs/0701072* (January 2007).

[221] Wang, C., Han, J., Jia, Y., Tang, J., Zhang, D., Yu, Y., and Guo, J. Mining Advisor-advisee Relationships from Research Publication Networks. In *Proceedings of the 16th International ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (Washington, District of Columbia, USA, July 25-28 2010), KDD '10, ACM, pp. 203–212.

[222] Wasserman, S., and Faust, K. *Social Network Analysis: Methods and Applications.* Cambridge University Press, Cambridge, UK, 1994.

[223] Whittaker, S., Jones, Q., Nardi, B., Creech, M., Terveen, L., Isaacs, E., and Hainsworth, J. ContactMap: Organizing Communication in a Social Desktop. *ACM Transactions on Computer-Human Interaction 11*, 4 (Dec. 2004), 445–471.

[224] Wilson, C., Boe, B., Sala, A., Puttaswamy, K. P., and Zhao, B. Y. User Interactions in Social Networks and Their Implications. In *Proceedings of the 4th European ACM Conference on Computer Systems* (Nuremberg, Germany, March 31-April 3 2009), EuroSys '09, ACM, pp. 205–218.

[225] Wu, A., DiMicco, J. M., and Millen, D. R. Detecting Professional Versus Personal Closeness Using an Enterprise Social Network Site. In *Proceedings of the 28th ACM SIGCHI International Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA, April 10-15 2010), CHI '10, ACM, pp. 1955–1964.

[226] Wu, S., Hofman, J. M., Mason, W. A., and Watts, D. J. Who Says What To Whom on Twitter. In *Proceedings of the 20th International Conference on World Wide Web* (Hyderabad, India, March 28-April 1 2011), WWW '11, ACM, pp. 705–714.

[227] Wu, T.-F., Lin, C.-J., and Weng, R. C. Probability Estimates for Multi-class Classification by Pairwise Coupling. *Journal of Machine Learning Research 5* (Dec. 2004), 975–1005.

[228] Xiang, R., Neville, J., and Rogati, M. Modeling Relationship Strength in Online Social Networks. In *Proceedings of the 19th International Conference on World Wide Web* (Raleigh, North Carolina, USA, April 26-30 2010), WWW '10, ACM, pp. 981–990.

[229] Xu, Z., Fu, Y., Mao, J., and Su, D. Towards the Semantic Web: Collaborative Tag Suggestions. In *Proceedings of the WWW Workshop on Collaborative Web Tagging* (Edinburgh, Scotland, UK, May 22 2006).

[230] Yamamoto, H., and Matsumura, N. Optimal Heterophily for Word-of-Mouth Diffusion. In *Proceedings of the 3rd AAAI International Conference on Weblogs and Social Media* (San Jose, California, USA, May 17-20 2009), ICWSM '09, AAAI Press, pp. 350–353.

[231] Yanbe, Y., Jatowt, A., Nakamura, S., and Tanaka, K. Can Social Bookmarking Enhance Search in the Web? In *Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries* (Vancouver, British Columbia, Canada, June 18-23 2007), JCDL '07, ACM, pp. 107–116.

[232] Yang, S.-H., Long, B., Smola, A., Sadagopan, N., Zheng, Z., and Zha, H. Like Like Alike: Joint Friendship and Interest Propagation in Social Networks. In *Proceedings of the 20th International Conference on World Wide Web* (Hyderabad, India, March 28-April 1 2011), WWW '11, ACM, pp. 537–546.

[233] Yang, Y.-H., Lin, Y.-C., Su, Y.-F., and Chen, H. H. A Regression Approach to Music Emotion Recognition. *IEEE Transactions on Audio, Speech & Language Processing 16*, 2 (February 2008), 448–457.

[234] Zanardi, V., and Capra, L. Social Ranking: Uncovering Relevant Content Using Tag-based Recommender Systems. In *Proceedings of the 2nd International ACM Conference on Recommender Systems* (Lausanne, Switzerland, October 23-25 2008), RecSys '08, ACM, pp. 51–58.

[235] Zerr, S., Bischoff, K., and Chernov, S. GuideMe! The World of Sights in Your Pocket. In *Proceedings of the 27th International IEEE Conference on Data Engineering* (Hannover, Germany, April 11-16 2011), ICDE '11, IEEE Computer Society, pp. 1348–1351.

[236] Zhang, L., Wu, X., and Yu, Y. Emergent Semantics from Folksonomies: A Quantitative Study. *Journal on Data Semantics VI: Special Issue on Emergent Semantics 4090* (2006), 168–186.

[237] Zheleva, E., and Getoor, L. To Join or Not to Join: The Illusion of Privacy in Social Networks with Mixed Public and Private User Profiles. In *Proceedings of the 18th International Conference on World Wide Web* (Madrid, Spain, April 20-24 2009), WWW '09, ACM, pp. 531–540.

[238] Zhou, D., Bian, J., Zheng, S., Zha, H., and Giles, C. L. Exploring Social Annotations for Information Retrieval. In *Proceedings of the 17th International Conference on World Wide Web* (Beijing, China, April 21-25 2008), WWW '08, ACM, pp. 715–724.

[239] Zollers, A. Emerging Motivations for Tagging: Expression, Performance, and Activism. In *Proceedings of the WWW Workshop on Tagging and Metadata for Social Information Organization* (Banff, Alberta, Canada, May 8 2007).

Curriculum Vitae

born on 1980/04/02 in Lauchhammer, Germany

Since Oct. 2006	**L3S Research Center / Leibniz Universität Hannover, Germany** Research associate / project manager and PhD student in computer science
Apr. 2006 – Sep. 2006	**University of Hildesheim, Germany** Graduate assistant at the Institute for Applied Linguistics - Faculty of Information and Communication Science
Oct. 2000 – Apr. 2006	**University of Hildesheim, Germany** Studies in International Information Management, Psychology, and Political Science (Magistra Artium)