Proceedings of International Joint Conference on Neural Networks, Atlanta, Georgia, USA, June 14-19, 2009

A Brief Survey of Computational Approaches in Social Computing

Irwin King, Jiexing Li, and Kam Tong Chan

***Abstract*— Web 2.0 technologies have brought new ways of connecting people in social networks for collaboration in various on-line communities. Social Computing is a novel and emerging computing paradigm that involves a multi-disciplinary approach in analyzing and modeling social behaviors on different media and platforms to produce intelligent and interactive applications and results. In this paper, we give a brief survey of the various machine learning and computational techniques used in Social Computing by first examining the social platforms, e.g., social network sites, social media, social games, social bookmarking, and social knowledge sites, where computational methodology is required to collect, extract, process, mine, and visualize the data. We then present surveys on more specific instances of computation tasks and techniques, e.g., social network analysis, link modeling and mining, ranking, sentiment analysis, etc., that are being used on these social platforms to obtain desirable results. Lastly, we present a small subset of an extensive reference list, which contains over 140 highly relevant references relating to the recent development in the computational aspects of Social Computing.**

Keywords: ***social computing, social networks, social media, collaborative filtering, social tagging, ranking, link analysis, graph mining, collaborative filtering, human computing, sentiment analysis***

I. Introduction

According to a new report from Netpop Research [1], 76% of all U.S. broadband users actively contribute to social media sites in one form or another, and 29% contribute regularly to social networking sites [2]. These social media sites include but not limited to Facebook, MySpace, YouTube, Flickr, iMeem, LastFM, Digg, Bebo, Google groups, hi5, LinkedIn, LiveJournal, etc. A quick lookup at Alexa [3] reveals that the the top five global sites that have the highest Internet traffic as of November 2008 are social networking or related sites as shown in Table. I. Hence, Social Computing is a social, cultural, as well as a computing phenomenon that cannot be ignored and begs for a more detailed investigation.

	USA	CHINA	GLOBAL
1	Google	Baidu	Yahoo
2	Yahoo	**QQ**	Google
3	**Myspace**	Sina	**YouTube**
4	**YouTube**	Google.cn	Windows Live
5	**Facebook**	Taobao	**Facebook**
6	Windows Live	163	MSN
7	MSN	Yahoo	**Myspace**
8	**Wikipedia**	Google	**Wikipedia**
9	EBay	Sohu	**Blogger**
10	AOL	Youku	Yahoo.jp

TABLE I

INTERNET TRAFFIC OF SOCIAL NETWORK SITES BY ALEXA AS OF NOVEMBER 2008

Social Computing is a novel and emerging computing paradigm that involves a multi-disciplinary approach in analyzing and modeling social behaviors on different media and platforms to produce intelligent applications [4], [5]. The multi-disciplinary underpinning includes computing, sociology, social psychology, organization theory, communication theory, human-computer interaction (HCI), etc. One of the better definitions on Social Computing by Wang et al. defined it as, "*Computational facilitation of social studies and human social dynamics as well as the design and use of information and communication technologies that consider social context.*" [5].

The three characteristics that capture the essence of Social Computing are:

- **Connectivity**–Here the emphasis on the relations among people within the group. Moreover, the medium of how they are connected could also be an important factor of how information is being transferred. These may include phone, email, instant messaging, SMS, chats, blogs, forums, social network services, and other emerging media [6].
- **Collaboration**–The way people collaborate is also important. Here, one can discuss ways that people can facilitate one another in a collaborative (positive) manner. Moreover, there are also adversarial or competitive (negative) relations among people that can affect adversely. Or, there could also be competitive collaborative learning [7]. Examples of these include collaborative filtering, trust and reputation systems, online auctions, verification games, social choices, knowledge sharing, etc.
- **Community**–The grouping or clustering of people is also another important factor in how we relate to one another. Communities may be formed through functional similarity, spatial closeness, or by other functional means. Communities become the collective source of wisdom. For example, these online communities are found in blogs, wikis, social networks, social tagging, collaborative filtering, collaborative bookmarking, podcasts, etc.

Although various computational techniques, e.g., classifi-

Irwin King, Jessie Li, and Thomas Chan are with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong (email: {king, jli, ktchan}@cse.cuhk.edu.hk).

This work is fully supported by two grants (Project No. CUHK 6902498 and Project No. CUHK 2050379) and is affiliated with the Microsoft-CUHK Joint Laboratory for Human-centric Computing and Interface Technologies.

cation, clustering, regression, etc. can be used in Social Computing, one major differentiating factor in the input data for Social Computing is the additional layer of personal and social contextual information. In other words, the input data is assumed to have latent information that one can take advantage when formulating results–it is no longer adequate to consider each record in the input space as independent, but rather linked with other input records. Therefore, intelligent computational approaches are needed since the information is dynamic, voluminous, nonlinear, and highly complex and the results need to be adaptive and intelligent. Hence, Social Computing has an added dimension to the traditional computational approaches.

The rest of this paper is organized as follows: Section II gives an overview of the various social phenomena that would require computational approaches. Section III introduces the various computational approaches in various Social Computing techniques. Finally, Section IV concludes the paper with some future views of Social Computing.

II. The Social Platforms

Here, we present several social platforms that give rise to social behaviors as input data for Social Computing.

A. Social Networks

In recent years, many social networking sites have sprung up as platforms to bring people together, e.g., FaceBook, MySpace, Xanga, QQ, Blogger, LinkedIn, hi5, etc. There are explicit links, e.g., direct links to friends, and also implicit links, e.g., participation in a discussion group, that give rise to added information for many mining applications ranging from community discovery to viral marketing. For example, [8] investigates the discovery of communities from communication documents collected temporally. It uses a constrained partitioning algorithm to separate the tripartite graphs of static communities at a specific time period based on topology and membership.

In the book [9], Barabasi illustrates a who-knows-who social network from general graph theoretic perspective. Ref. [10] provides a correlation between a person's social group and his/her personal behavior. A lot of research work has also been done for each specific topic. For instance, there are research about Blogger [11], [12], [13] and other social networking sites.

B. Social Media

Social media are primarily Internet and mobile based tools for sharing and discussing information among human beings. One of the recently emerging social platforms that is brought on by the ubiquitous presence of media capturing devices such as phones, digital cameras, video recorders, etc. They include social media sites such as Flickr, YouTube, etc.

Ref. [14] presents a general ranking framework for factual information retrieval from social media sites. The framework is effective at retrieving well-formed, factual answers and can be tuned with minimum of manual labeling. A robust ranking method was proposed in [15].

Ref. [16] devises a number of attributes of video users and their social behaviour from a large test collection of YouTube users which could potentially be used to detect spammers in video social media sites.

Ref. [17] provides a game for the annotation of music and sound. Extensive researches [8], [18], [19], [20], [21] have been focused on the retrieval of spoken documents, image, etc.

C. Social Games/Human Computation

Social games or Games with A Purpose (GWAP) [22], [23], [24] is an innovative idea that makes use of human brain power to solve difficult problems. The kind of problems that we are dealing with have two things in common: (1) they are problems that computers are not good at solving, and (2) they are trivial for humans. For example, image annotation is a task that tries to figure out the objects in an image by asking the player to description them. It is a task that is nearly impossible for computer to accomplish with high accuracy in a general setting. With social games, the image annotation problem can now be solved by a series of interaction between humans and the computers.

Social games are designed in such a way that people play them just because of fun [25]. People playing those games do not necessarily know that they are solving a meta problem. Take ESPGame [26] as an example, players joining the same game section have the same image displayed on their screen at the same time. They are asked to guess whatever the other player is typing instead of label the image directly. The players, while having no other common information except the given image, will try to name the objects inside the image, achieving the goal of image annotation. Players can have a sense of achievement if they can make a successful guess, making them more willing to player the game.

Various games deploy different interaction models. Since social games solve problem by people contributing the answers voluntary, we need mechanisms to ensure that the answers meet a certain quality standard. The interaction models play the most important role in achieving this quality guarantee. Existing games can be divided into three categories depending their the interaction models, namely, output-agreement games, inversion-problem games and input agreement games [25].

Output-Agreement Games have all the players playing the same role. Players are selected randomly to join a game session to ensure their anonymity. All of them are given the same input and are asked to generate outputs. An answer is said to have verified when their outputs are agree with each other. Examples of this type of game include the ESPGame [26], an image annotation engine and reCAPTCHA [27], a character recognition system.

Inversion-Problem Games divide players into two roles. Again, players are selected randomly to join a game session. In each round of the game, a player is chosen to act as a "describer", who has access to the input while the other players act as a "guesser". The describer has to give hints to the guesser so that they can guess the original input. Communication constraints are added so that describer cannot tell the answer to the guesser directly. Under this setting, if a guesser successfully guess the input, we say that the hints given by the describer are correct descriptions it. Examples of this type of game include Peekaboom [28], a system for locating objects in images, Phetch [29], an image description generator and [30], a common sense collection engine.

Input-Agreement Games are similar to output-agreement games in a sense that all players are assigned to the same role. Each player is given an input which may or may not be the same with other players. They are asked to guess whether they share the same input with the other player. They are allowed to share their descriptions to the input so as to make their guesses. An example of this type of game is [17].

D. Social Bookmarking/Tagging

Social bookmarking can be used for Internet users to store, search, and manage bookmarks of web pages on the Internet. Social bookmarking and tagging sites such as Del.icio.us, StumbleUpon, Ma.gnolia, etc. have helped search engines to index sites faster and give more quality results by analyzing the inputs about a site from the users. However, one of the problems in social tagging is that different people express the same concept differently. This often leads to low precision retrieval due to the potentially large number of synonyms exist in the system. Ref. [31] utilizes user preference profiles to identify synonyms that can be used to retrieve more relevant documents through user's query expansion.

The underlying data structures of Social bookmarking, known as folksonomies, which consist of large-scale bodies of lightweight annotations, were studied in [32], [33], [34]. A clustering approach for computing such a conceptual hierarchy for a given folksonomy was presented in [35].

Difference aspects of the social bookmark technique have been studied. Variety search methods [31], [36], [37], [38], [39] utilizing social bookmarking were developed, which aimed at improving searching accuracy or efficiency. Other researches [40], [41], [42], [43], [44] presented analyses of the structure or the organization of bookmark systems.

In [45], authors investigated the social tag prediction problem, which tried to predict whether a given tag could/should be applied to a particular object, while [46] introduced an approach to eliminate spam in social bookmarking systems using machine learning approaches.

E. Social News and Social Knowledge Sharing

Social news refers to websites where users submit their own information. Users can also vote on links to determine which links are presented. The collective wisdom is elegantly manifested in the Wikipedia site where it is being used by millions of people and edited by thousands of people everyday. The notion that each individual contributes to a collective pool of knowledge is further being realized in automated Question Answering (QA) systems such as AnswerBus, Webclopedia, Yahoo's babelfish, etc. Ref. [47] presented a content-driven reputation system for Wikipedia authors, in which authors gain reputation when their edits to Wikipedia articles are preserved by subsequent authors, and they lose reputation when their edits are rolled back. Ref. [48] identified several types of structures which can be automatically enhanced in Wikipedia.

Since there are huge amount of articles available on the web, filtering algorithms of netnews [49], [50], [51] were introduced to help people find articles they would like.

III. Computational Tasks and Techniques in Social Computing

A. Social Network Theory, Modeling, and Analysis

Characterizing the relationship that exists between a person's social group and his/her personal behavior has been a long standing goal of social network analysis. The seminal empirical study of the structure of social networks was done by Michael Gurevich in 1961 [52]. Subsequently, Stanley Milgram continued the work in acquaintanceship networks, which led to his work in the Small World Problem [53], [54].

Not surprisingly, all networks are not so chaotic or random as once assumed, but rather they have underlying structures and follow simple rules. Understanding the structure of these networks will allow us to gain an insight to design better network structures for an organization, stop the spread of viruses, select how to propagate information effectively, check the robustness of the network, etc. [55], [56], [57].

Park [6] identifies the hyperlink network among websites as an emerging computational methodology that can describe social behaviors of social agents on the web.

The book "Models and Methods in Social Network Analysis" [58] presents the most important developments in quantitative models and methods for analyzing social network data. Different models for social computing were proposed in [59], [60]. New challenges for social computing were also analyzed in [61].

B. Ranking

In any information retrieval system, ranking of the retrieved results is a crucial task of paramount importance. A good ranking scheme will give highly relevant and accurate results requiring minimum amount of computational resources. There is a wealth of work on this topic and we only highlight a few here.

Ref. [11] proposed a way to integrate an opinion identification toolkit into the retrieval process of an Information Retrieval system, such that opinionated, relevant documents are retrieved in response to a query. Ref. [62] conducted a study on the approach of directly optimizing evaluation measures in learning to rank for Information Retrieval. Algorithms or applications of ranking were studies in [63], [64], [65], [66], [67], [68].

If we divide the algorithms according to the targets they are going to rank, we have the following categories. Algorithms for graph ranking were described in [69], [70], [71]. Ranking methods for mediators were presented in [72]. Variations of Pagerank methods were development in [73],

[74], [75]. Ranking algorithms for objects [76], [77] and for data lying in the Euclidean space, such as text or image data [78] were also proposed.

C. Query Log Processing

Search engines as well as social network sites collect a voluminous amount of query log or click-through data from their users. This is a gold mine of information that can be used to improve retrieval results

By mining web click-through data, [79] proposes a method to automatically acquire query translation pairs. Ref. [80] developed a two-level query suggestion model by mining clickthrough data which provides semantically relevant queries for users.

It was observed that users searching the web often perform a sequence, or chain, of queries with a similar information need. Motivated by this, [81] used clickthrough data to learn ranked retrieval functions for web search results.

Ref. [82] presented an approach to automatically optimizing the retrieval quality of search engines, which utilized clickthrough data for training, namely the query-log of the search engine in connection with the log of links.

D. Web Spam Detection

Ref. [83] presented an algorithm, witch, that learns to detect spam hosts or pages on the Web. Unlike most other approaches, it simultaneously exploits the structure of the Web graph as well as page contents and features.

Different algorithms have been developed for the detection of some particular type of spam. In [16], method for identifying video spam was proposed. Ref. [46] aimed at detecting spam in social bookmarking systems. Some focus on link or web page spam detection [84], [85].

Motivated by the heat diffusion phenomena, a DiffusionRank [86] algorithm was developed as a possible penicillin for web spam. Ref. [87] proposed techniques to semi-automatically separate reputable, good pages from spam by combating web spam with TrustRank. The algorithm first selected a small set of seed pages to be evaluated by an expert. Then link structure of the web was used to discover other pages that are likely to be good. Some other methods were also proposed in [88], [89].

In [15], authors outline a machine learning- based ranking framework for social media that integrates user interactions and content relevance, and demonstrate its effectiveness for answer retrieval in a popular community question answering portal. Moreover, the work describes a vote spam attack model that will make the framework more robust.

Ref. [19] uses content and other relevant information such as links between items and quality ratings from members of the community to automatically identify high quality content, particularly on Yahoo! Answers.

E. Graph/Link Analysis and Mining

As social relations can be modeled using graphs, link analysis and modeling is then a natural way to process these social graphs. Here, we often are interested in the qualitative and quantitative measurements of the graphs using links. In particular, the desirable graph mining algorithms would have the following characteristics:

1) **Efficiency**: Evidently, only those algorithms with low time and space complexity are practical and applicable to the immense size of the Web.
2) **Scalability**: The dramatic growth rate of the Web poses a serious challenge of scalability for web applications that aspire to cover a large part of the Web.
3) **Stability**: An algorithm should be stable to perturbations of the Web, including link structure and content of web pages.
4) **Robustness**: We use the term robust to indicate that an algorithm on theWeb is resistent to commonly used web spamming techniques.

There is already a slew of excellent surveys on link and graph analysis [6], [90], [91], [92] and link mining [93], [94], [95], [96].

Moreover, there is also a wealth of information on graph mining [97], [98], [99], e.g., community structure, density, centrality, centralization, components, cores, cliques, cycles, knots, positions, roles, clusters, etc.

The link structure of web can be used to improve the performance of other algorithms. Ref. [100] described a two-stage approach which uses link structure to improve Web Spam Classifiers. Ref. [101], [52] showed how to exploit the web link graph structure to speed up the computation of PageRank.

F. Collaborative Filtering

Collaborative filtering is the process of identifying information interest of a specific user based on the information provided by other similar users. In the following, we give a brief overview of the many collaborative filtering algorithms that have been developed.

1. Memory-based collaborative filtering. The techniques are also known as nearest-neighbor collaborative filtering. The most common form of neighborhood-based approach is the user-based model [102], [103], [49]. A framework for performing collaborative filtering was presented in [104]. Ref. [105] divided the neighborhood-based prediction approach into three components identified as similarity computation, neighbor selection, and rating combination. GroupLens [50] is a system for collaborative filtering of netnews. An alternative form of the neighborhood-based approach is the item-based model [106], [107], [108]. It shares the same idea with user-based method. Amazon.com [106] used item-to-item collaborative filtering to personalize the online store for each customer. Ref. [107] looked into different techniques for computing item-item similarity.

2. Model-based Collaborative Filtering Algorithms. In the model-based approaches, a predefined model of user ratings is developed by using training datasets.

Common used model-based approaches are clustering models [109], [110]. Ref. [111] presented an algorithm for collaborative filtering based on hierarchical clustering.

Algorithms may also use latent factor model for collaborative filtering. In [112], a peer-to-peer protocol for collaborative filtering was proposed to protect the privacy of individual data. Its algorithm was based on factor analysis which has advantages in speed and storage over previous algorithms. Aspect method model is a probabilistic latent semantic model, which models users preferences as a convex combination of preference factors [113]. Previous works in aspect method model include [114], [115]. Authors in [116] proposed an algorithm based on a generalization of probabilistic latent semantic analysis to continuous valued response variables.

The model-based approaches are often time-consuming to build and update, and cannot cover as diverse a user range as the memory-based approaches do.

3. Other Related Approaches. Hybrid frameworks [117], [51] and attack resistant collaborative algorithms were also introduced.

To address the problem of sparsity inherent to rating data, [108] reformulated the memory collaborative filtering problem in a generative probability framework. Ref. [118] proposed an effective missing data prediction algorithm, in which information of both users and items is taken into account. The work was further expanded by using probabilistic matrix factorization for social recommendation [119]. Ref. [120] described a new method called personality diagnosis which retains some of the advantages of traditional similarity-weight techniques and in addition, has a meaningful probability interpretation. Ref. [110] introduced a smoothing-based method which combined the advantages of memory-base and model-based approaches. As pointed out in [121], [122], some users may faithfully express their true opinion. These noisy or incorrect ratings can tamper the quality of the recommendation systems. As a result, it is necessary for systems to provide guarantees on the robustness of recommendations to ensure continued user trust.

Some other algorithms have been proposed [123], [124]. Ref. [125] introduced a collaborative filtering approach that addresses item ranking problem directly by modeling user preferences derived from the ratings. At the same time, in order to solve the data sparsity problem, researchers proposed dimensionality reduction approaches in [126].

G. Sentiment Analysis and Opinion Mining

Opinion mining focuses on extracting people's opinion from the web. The recent expansion of the web encourages users to contribute and express themselves via blogs, videos, social networking sites, etc. All these provide a huge amount of valuable information that we are interested to analyze. Given a piece of text, opinion mining systems analyze 1) which part is opinion expressing, 2) who wrote the opinion, 3) what is being commented on and 4) what is the opinion of the writer. Many previous works have been proposed to solve various problems in opinion mining:

1) Feature Extraction: By feature extraction, we mean that given a text document, we try to extract the target object's feature commented by the author. Hu [127] proposed an approach based on association rule mining to perform feature extraction. This is based on the observation that features are frequently mentioned in the reviews. Yi [128] proposed to incorporate NLP (Natural Language Processing) techniques into the feature extraction process. Based on the part-of-speech (POS) tag of the review text, they extract terms matching a predefined set of patterns. Then they use statistical techniques to prune away the non-feature terms.

2) Sentiment Analysis: Sentiment analysis aims at determining the polarity (i.e., positive or negative) of a piece of text. Early works in this area are mostly based on lexicon based approaches [127], [129], [130], [131]. The idea is to build a lexicon of words with known sentiment for sentiment classification. Pang [132] studied the performance of using traditional machine learning techniques to perform sentiment analysis in document level. Turney [133] proposed an unsupervised sentiment classification method based on Pointwise Mutual Information (PMI) between the words.

IV. Conclusion

We have presented a brief overview on the computational aspects of Social Computing. We have examined the social media and platforms that are being used, e.g., Social Networks, Social Bookmarking, Social Tagging, Social Media, etc. where social behaviors can be observed and collected as data for further processing. Moreover, we also survey computational tasks and approaches that are being used on these platforms. We observe that the latent information among people in communities give rise to the exciting prospect to view and process social computation differently than what we have done before.

This survey is only the tip of the iceberg in Social Computing as we are finding more ways people connect, collaborate, and form communities on the Web. We plan to expand the survey by providing more detailed and recent work done in Social Computing in the future.

References

[1] "Netpop research." [Online]. Available: http://www.netpopresearch.com/

[2] "Forty million americans now contribute to social networking sites: Who are they?" [Online]. Available: http://www.readwriteweb.com/archives/forty_million_american_now_con.php

[3] "Alexa: The web information company." [Online]. Available: http://www.alexa.com/

[4] F.-Y. Wang, "Toward a paradigm shift in social computing: The acp approach," *IEEE Intelligent Systems*, vol. 22, no. 5, pp. 65–67, 2007.

[5] F.-Y. Wang, K. M. Carley, D. Zeng, and W. Mao, "Social computing: From social informatics to social intelligence," *IEEE Intelligent Systems*, vol. 22, no. 2, pp. 79–83, 2007.

[6] H. W. Park, "Hyperlink network analysis: A new method for the study of social structure on the web," *Connections*, vol. 25, no. 1, pp. 46–61, 2003.

[7] B. Awerbuch and R. D. Kleinberg, "Competitive collaborative learning," in *COLT*, ser. Lecture Notes in Computer Science, P. Auer and R. Meir, Eds., vol. 3559. Springer, 2005, pp. 233–248.

[8] D. Zhou, I. G. Councill, H. Zha, and C. L. Giles, "Discovering temporal communities from social network documents," in *ICDM*. IEEE Computer Society, 2007, pp. 745–750.

[9] A.-L. Barabsi, *Linked: The New Science of Networks*. Cambridge, MA: Perseus Publishing, 2002.

[10] P. Singla and M. Richardson, "Yes, there is a correlation: - from social networks to personal behavior on the web," in *WWW*, J. Huai, R. Chen, H.-W. Hon, Y. Liu, W.-Y. Ma, A. Tomkins, and X. Zhang, Eds. ACM, 2008, pp. 655–664.

[11] B. He, C. Macdonald, and I. Ounis, "Ranking opinionated blog posts using opinionfinder," in *SIGIR*, S.-H. Myaeng, D. W. Oard, F. Sebastiani, T.-S. Chua, and M.-K. Leong, Eds. ACM, 2008, pp. 727–728.

[12] A. Sun, M. Hu, and E.-P. Lim, "Searching blogs and news: a study on popular queries," in *SIGIR*, S.-H. Myaeng, D. W. Oard, F. Sebastiani, T.-S. Chua, and M.-K. Leong, Eds. ACM, 2008, pp. 729–730.

[13] J. L. Elsas, J. Arguello, J. Callan, and J. G. Carbonell, "Retrieval and feedback models for blog feed search," in *SIGIR*, S.-H. Myaeng, D. W. Oard, F. Sebastiani, T.-S. Chua, and M.-K. Leong, Eds. ACM, 2008, pp. 347–354.

[14] J. Bian, Y. Liu, E. Agichtein, and H. Zha, "Finding the right facts in the crowd: factoid question answering over social media," in *WWW*, J. Huai, R. Chen, H.-W. Hon, Y. Liu, W.-Y. Ma, A. Tomkins, and X. Zhang, Eds. ACM, 2008, pp. 467–476.

[15] ——, "A few bad votes too many?: towards robust ranking in social media," in *AIRWeb*, ser. ACM International Conference Proceeding Series, C. Castillo, K. Chellapilla, and D. Fetterly, Eds., 2008, pp. 53–60.

[16] F. Benevenuto, T. Rodrigues, V. Almeida, J. M. Almeida, C. Zhang, and K. Ross, "Identifying video spammers in online social networks," in *AIRWeb*, ser. ACM International Conference Proceeding Series, C. Castillo, K. Chellapilla, and D. Fetterly, Eds., 2008, pp. 45–52.

[17] R. B. D. M. C. Edith L. M. Law, Luis von Ahn, "Tagatune: A game for music and sound annotation," *ISMIP*, 2007.

[18] T. K. Chia, K. C. Sim, H. Li, and H. T. Ng, "A lattice-based approach to query-by-example spoken document retrieval," in *SIGIR*, S.-H. Myaeng, D. W. Oard, F. Sebastiani, T.-S. Chua, and M.-K. Leong, Eds. ACM, 2008, pp. 363–370.

[19] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne, "Finding high-quality content in social media," in *WSDM*, M. Najork, A. Z. Broder, and S. Chakrabarti, Eds. ACM, 2008, pp. 183–194.

[20] C. Wang, L. Zhang, and H.-J. Zhang, "Learning to reduce the semantic gap in web image retrieval and annotation," in *SIGIR*, S.-H. Myaeng, D. W. Oard, F. Sebastiani, T.-S. Chua, and M.-K. Leong, Eds. ACM, 2008, pp. 355–362.

[21] G. Luo and C. Tang, "On iterative intelligent medical search," in *SIGIR*, S.-H. Myaeng, D. W. Oard, F. Sebastiani, T.-S. Chua, and M.-K. Leong, Eds. ACM, 2008, pp. 3–10.

[22] L. von Ahn, "Games with a purpose," *IEEE Computer*, vol. 39, no. 6, pp. 92–94, 2006.

[23] C.-W. Lin, K. T. Chen, L. J. Chen, I. King, and J. H.-M. Lee, "An analytical approach to optimizing the utility of esp games," in *Proceedings to the IEEE/WIC/ACM WI-2008 (WI2008)*, Sydney, Australia, December, 2008.

[24] L. J. Chen, B.-C. Wang, K. T. Chen, I. King, and J. H.-M. Lee, "An analytical study of puzzle selection strategies for the esp game," in *Proceedings to the IEEE/WIC/ACM WI-2008 (WI2008)*, Sydney, Australia, December, 2008.

[25] L. von Ahn and L. Dabbish, "Designing games with a purpose," *Commun. ACM*, vol. 51, no. 8, pp. 58–67, 2008. [Online]. Available: http://dblp.uni-trier.de/db/journals/cacm/cacm51.html#AhnD08

[26] ——, "Labeling images with a computer game," in *CHI*, E. Dykstra-Erickson and M. Tscheligi, Eds. ACM, 2004, pp. 319–326.

[27] L. von Ahn, B. Maurer, C. McMillen, D. Abraham, and M. Blum, "reCAPTCHA: Human-Based Character Recognition via Web Security Measures," *Science*, vol. 321, no. 5895, p. 1465, 2008.

[28] L. von Ahn, R. Liu, and M. Blum, "Peekaboom: a game for locating objects in images," in *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*. New York, NY, USA: ACM, 2006, pp. 55–64.

[29] L. von Ahn, S. Ginosar, M. Kedia, and M. Blum, "Improving Image Search with PHETCH," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 4, 2007.

[30] L. von Ahn, M. Kedia, and M. Blum, "Verbosity: a game for collecting common-sense facts," in *CHI*, R. E. Grinter, T. Rodden, P. M. Aoki, E. Cutrell, R. Jeffries, and G. M. Olson, Eds. ACM, 2006, pp. 75–78.

[31] M. Clements, A. P. de Vries, and M. J. T. Reinders, "Detecting synonyms in social tagging systems to improve content retrieval," in *SIGIR*, S.-H. Myaeng, D. W. Oard, F. Sebastiani, T.-S. Chua, and M.-K. Leong, Eds. ACM, 2008, pp. 739–740.

[32] R. Jäschke, L. B. Marinho, A. Hotho, L. Schmidt-Thieme, and G. Stumme, "Tag recommendations in folksonomies," in *LWA*, A. Hinneburg, Ed. Martin-Luther-University Halle-Wittenberg, 2007, pp. 13–20.

[33] A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme, "Information retrieval in folksonomies: Search and ranking," in *ESWC*, 2006, pp. 411–426.

[34] C. Cattuto, D. Benz, A. Hotho, and G. Stumme, "Semantic grounding of tag relatedness in social bookmarking systems," in *International Semantic Web Conference*, 2008, pp. 615–631.

[35] M. Grahl, A. Hotho, and G. Stumme, "Conceptual clustering of social bookmarking sites," in *LWA*, A. Hinneburg, Ed. Martin-Luther-University Halle-Wittenberg, 2007, pp. 50–54.

[36] R. Schenkel, T. Crecelius, M. Kacimi, S. Michel, T. Neumann, J. X. Parreira, and G. Weikum, "Efficient top-k querying over social-tagging networks," in *SIGIR*, S.-H. Myaeng, D. W. Oard, F. Sebastiani, T.-S. Chua, and M.-K. Leong, Eds. ACM, 2008, pp. 523–530.

[37] M. Okamoto, M. Kikuchi, and T. Yamasaki, "One-button search extracts wider interests: an empirical study with video bookmarking search," in *SIGIR*, S.-H. Myaeng, D. W. Oard, F. Sebastiani, T.-S. Chua, and M.-K. Leong, Eds. ACM, 2008, pp. 779–780.

[38] P. Heymann, G. Koutrika, and H. Garcia-Molina, "Can social bookmarking improve web search?" in *WSDM*, M. Najork, A. Z. Broder, and S. Chakrabarti, Eds. ACM, 2008, pp. 195–206.

[39] B. Krause, A. Hotho, and G. Stumme, "A comparison of social bookmarking with traditional search," in *ECIR*, 2008, pp. 101–113.

[40] G. W. Furnas, C. Fake, L. von Ahn, J. Schachter, S. A. Golder, K. Fox, M. Davis, C. Marlow, and M. Naaman, "Why do tagging systems work?" in *CHI Extended Abstracts*, G. M. Olson and R. Jeffries, Eds. ACM, 2006, pp. 36–39.

[41] *BibSonomy: A Social Bookmark and Publication Sharing System*. Aalborg University Press, 2006.

[42] S. A. Golder and B. A. Huberman, "The structure of collaborative tagging systems," *CoRR*, vol. abs/cs/0508082, 2005.

[43] R. Jäschke, M. Grahl, A. Hotho, B. Krause, C. Schmitz, and G. Stumme, "Organizing publications and bookmarks in bibsonomy," in *CKC*, 2007.

[44] J. Park, T. Fukuhara, I. Ohmukai, H. Takeda, and S. goo Lee, "Web content summarization using social bookmarks: a new approach for social summarization," in *WIDM '08: Proceeding of the 10th ACM workshop on Web information and data management*. New York, NY, USA: ACM, 2008, pp. 103–110.

[45] P. Heymann, D. Ramage, and H. Garcia-Molina, "Social tag prediction," in *SIGIR*, S.-H. Myaeng, D. W. Oard, F. Sebastiani, T.-S. Chua, and M.-K. Leong, Eds. ACM, 2008, pp. 531–538.

[46] B. Krause, C. Schmitz, A. Hotho, and G. Stumme, "The anti-social tagger: detecting spam in social bookmarking systems," in *AIRWeb*, ser. ACM International Conference Proceeding Series, C. Castillo, K. Chellapilla, and D. Fetterly, Eds., 2008, pp. 61–68.

[47] B. T. Adler and L. de Alfaro, "A content-driven reputation system for the wikipedia," in *WWW*, 2007, pp. 261–270.

[48] F. Wu and D. S. Weld, "Autonomously semantifying wikipedia," in *CIKM '07: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*. New York, NY, USA: ACM, 2007, pp. 41–50.

[49] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl, "Grouplens: Applying collaborative filtering to usenet news," *Commun. ACM*, vol. 40, no. 3, pp. 77–87, 1997.

[50] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "Grouplens: An open architecture for collaborative filtering of netnews," in *CSCW*, 1994, pp. 175–186.

[51] *Combining content-based and collaborative filters in an online newspaper*, 1999.

[52] M. Gurevich, *The Social Structure of Acquaintanceship Netowrks*. MIT Press, Cambridge, MA, 1961.

[53] S. Milgram, "The small world problem," *Psychology Today*, vol. 2, pp. 60–67, 1967.

[54] ——, "An experimental study of the small world problem," *Sociometry*, vol. 32, pp. 425–443, 1969.

[55] D. J. W. Mark Newman, Albert-Laszlo Barabasi, *The Structure and Dynamics of Networks*. Princeton University Press, 2006.

[56] D. J. Watts, *Six Degrees: The Science of a Connected Age*. W. W. Norton and Company, 2004.

[57] A.-L. Barabási, "Linked: The new science of networks," *J. Artificial Societies and Social Simulation*, vol. 6, no. 2, 2003.

[58] S. W. Peter J. Carrington, John Scott, *Models and methods in social network analysis*. Cambridge University Press, 2005.
[59] X. Li, W. Mao, D. Zeng, and F.-Y. Wang, "Agent-based social simulation and modeling in social computing," in *ISI Workshops*, ser. Lecture Notes in Computer Science, C. C. Yang, H. Chen, M. Chau, K. Chang, S.-D. Lang, P. S. Chen, R. Hsieh, D. Zeng, F.-Y. Wang, K. M. Carley, W. Mao, and J. Zhan, Eds., vol. 5075. Springer, 2008, pp. 401–412.
[60] J. M. Kleinberg, "Distributed social systems," in *PODC*, E. Ruppert and D. Malkhi, Eds. ACM, 2006, pp. 5–6.
[61] T. Yamakami, "Ubiquitous computing literacy: A new challenge for social computing," in *EUC*, ser. Lecture Notes in Computer Science, L. T. Yang, M. Guo, G. R. Gao, and N. K. Jha, Eds., vol. 3207. Springer, 2004, pp. 712–723.
[62] J. Xu, T.-Y. Liu, M. Lu, H. Li, and W.-Y. Ma, "Directly optimizing evaluation measures in learning to rank," in *SIGIR*, S.-H. Myaeng, D. W. Oard, F. Sebastiani, T.-S. Chua, and M.-K. Leong, Eds. ACM, 2008, pp. 107–114.
[63] K. Zhou, G.-R. Xue, H. Zha, and Y. Yu, "Learning to rank with ties," in *SIGIR*, S.-H. Myaeng, D. W. Oard, F. Sebastiani, T.-S. Chua, and M.-K. Leong, Eds. ACM, 2008, pp. 275–282.
[64] R. Jin, H. Valizadegan, and H. Li, "Ranking refinement and its application to information retrieval," in *WWW*, J. Huai, R. Chen, H.-W. Hon, Y. Liu, W.-Y. Ma, A. Tomkins, and X. Zhang, Eds. ACM, 2008, pp. 397–406.
[65] M.-R. Amini, T.-V. Truong, and C. Goutte, "A boosting algorithm for learning bipartite ranking functions with partially labeled data," in *SIGIR*, S.-H. Myaeng, D. W. Oard, F. Sebastiani, T.-S. Chua, and M.-K. Leong, Eds. ACM, 2008, pp. 99–106.
[66] C. J. C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. N. Hullender, "Learning to rank using gradient descent," in *ICML*, ser. ACM International Conference Proceeding Series, L. D. Raedt and S. Wrobel, Eds., vol. 119. ACM, 2005, pp. 89–96.
[67] N. Eiron, K. S. McCurley, and J. A. Tomlin, "Ranking the web frontier," in *WWW*, S. I. Feldman, M. Uretsky, M. Najork, and C. E. Wills, Eds. ACM, 2004, pp. 309–318.
[68] H. Yang, I. King, and M. R. Lyu, "Predictive ranking: A novel page ranking approach by estimating the web structure," in *Poster Proceedings of the Fourteenth International World Wide Web Conference*, A. Ellis and T. Hagino, Eds., Keio University. ACM, May 10-14, 2005, pp. 944–945. [Online]. Available: http://www2005.org/cdrom/docs/p944.pdf
[69] ——, "Predictive random graph ranking on the web," in *2006 International Joint Conference on Neural Networks (IJCNN2006)*, Vancouver, Canada, July 16-21, 2006, pp. 1825–1832. [Online]. Available: http://www.cse.cuhk.edu.hk/∼king/PUB/IJCNN2006_Yang.pdf
[70] H. Deng, I. King, and M. R. Lyu, "Effective latent space graph-based re-ranking model with global consistency," in *Proceedings to the Second ACM International Conference on Web Search and Data Mining (WSDM2009)*, Barcelona, Spain, February 9-12, 2009.
[71] A. Agarwal and S. Chakrabarti, "Learning random walks to rank nodes in graphs," in *ICML*, ser. ACM International Conference Proceeding Series, Z. Ghahramani, Ed., vol. 227. ACM, 2007, pp. 9–16.
[72] R. Saga and H. Tsuji, "Ranking method for mediators in social network," in *HCI (15)*, ser. Lecture Notes in Computer Science, D. Schuler, Ed., vol. 4564. Springer, 2007, pp. 151–159.
[73] R. M. L. Page, S. Brin and T. Winograd, "The pagerank citation ranking: Bringing order to the web," *Stanford Digital Library Technologies Project*, 1999.
[74] O. Kurland and L. Lee, "Pagerank without hyperlinks: Structural re-ranking using links induced by language models," *CoRR*, vol. abs/cs/0601045, 2006.
[75] R. A. Baeza-Yates, P. Boldi, and C. Castillo, "Generalizing pagerank: damping functions for link-based ranking algorithms," in *SIGIR*, 2006, pp. 308–315.
[76] Z. Nie, Y. Zhang, J.-R. Wen, and W.-Y. Ma, "Object-level ranking: bringing order to web objects," in *WWW*, A. Ellis and T. Hagino, Eds. ACM, 2005, pp. 567–574.
[77] T. Qin, T.-Y. Liu, X.-D. Zhang, D.-S. Wang, W.-Y. Xiong, and H. Li, "Learning to rank relational objects and its application to web search," in *WWW*, J. Huai, R. Chen, H.-W. Hon, Y. Liu, W.-Y. Ma, A. Tomkins, and X. Zhang, Eds. ACM, 2008, pp. 407–416.
[78] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Schölkopf, "Ranking on data manifolds," in *NIPS*, S. Thrun, L. K. Saul, and B. Schölkopf, Eds. MIT Press, 2003.
[79] R. Hu, W. Chen, P. Bai, Y. Lu, Z. Chen, and Q. Yang, "Web query translation via web log mining," in *SIGIR*, S.-H. Myaeng, D. W. Oard, F. Sebastiani, T.-S. Chua, and M.-K. Leong, Eds. ACM, 2008, pp. 749–750.
[80] H. Ma, H. Yang, I. King, and M. R. Lyu, "Learning latent semantic relations from clickthrough data for query suggestion," in *CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge management*. New York, NY, USA: ACM, 2008, pp. 709–718.
[81] F. Radlinski and T. Joachims, "Query chains: Learning to rank from implicit feedback," *CoRR*, vol. abs/cs/0605035, 2006.
[82] T. Joachims, "Optimizing search engines using clickthrough data," in *KDD '02: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM, 2002, pp. 133–142.
[83] J. Abernethy, O. Chapelle, and C. Castillo, "Web spam identification through content and hyperlinks," in *AIRWeb*, ser. ACM International Conference Proceeding Series, C. Castillo, K. Chellapilla, and D. Fetterly, Eds., 2008, pp. 41–44.
[84] A. Ntoulas, M. Najork, M. Manasse, and D. Fetterly, "Detecting spam web pages through content analysis," in *WWW*, L. Carr, D. D. Roure, A. Iyengar, C. A. Goble, and M. Dahlin, Eds. ACM, 2006, pp. 83–92.
[85] D. Zhou, C. Burges, and T. Tao, "Transductive link spam detection," in *AIRWeb*, ser. ACM International Conference Proceeding Series, vol. 215, 2007.
[86] H. Yang, I. King, and M. R. Lyu, "Diffusionrank: A possible penicillin for web spamming," in *Proceedings to the Special Interest Group on Information Retrieval (SIGIR2007)*, N. Kando, C. Clarke, and N. Fuhr, Eds. Amsterdam, Netherlands: ACM, 23-27 July 2007, pp. 431–438. [Online]. Available: http://www.cse.cuhk.edu.hk/∼king/PUB/SIGIR07_Yang.pdf
[87] Z. Gyöngyi, H. Garcia-Molina, and J. O. Pedersen, "Combating web spam with trustrank," in *VLDB*, M. A. Nascimento, M. T. Özsu, D. Kossmann, R. J. Miller, J. A. Blakeley, and K. B. Schiefer, Eds. Morgan Kaufmann, 2004, pp. 576–587.
[88] R. Andersen, C. Borgs, J. T. Chayes, J. E. Hopcroft, K. Jain, V. S. Mirrokni, and S.-H. Teng, "Robust pagerank and locally computable spam detection features," in *AIRWeb*, ser. ACM International Conference Proceeding Series, C. Castillo, K. Chellapilla, and D. Fetterly, Eds., 2008, pp. 69–76.
[89] B. Mehta, T. Hofmann, and P. Fankhauser, "Lies and propaganda: detecting spam users in collaborative filtering," in *Intelligent User Interfaces*, D. N. Chin, M. X. Zhou, T. A. Lau, and A. R. Puerta, Eds. ACM, 2007, pp. 14–21.
[90] J. Scott, *Social network analysis: a handbook*. SAGE Publications, 2000.
[91] J. Kleinberg, S. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, "The Web as a Graph: Measurements, Models and Methods," *Lecture Notes in Computer Science*, pp. 1–17, 1999.
[92] A. Borodin, G. Roberts, J. Rosenthal, and P. Tsaparas, "Link analysis ranking: algorithms, theory, and experiments," *ACM Transactions on Internet Technology (TOIT)*, vol. 5, no. 1, pp. 231–297, 2005.
[93] S. Chakrabarti, B. Dom, D. Gibson, J. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, "Mining the Link Structure of the World Wide Web," *IEEE Computer*, vol. 32, no. 8, pp. 60–67, 1999.
[94] R. Kosala and H. Blockeel, "Web mining research: a survey," *ACM SIGKDD Explorations Newsletter*, vol. 2, no. 1, pp. 1–15, 2000.
[95] L. Getoor and C. Diehl, "Link mining: a survey," *SIGKDD Explorations*, vol. 7, no. 2, pp. 3–12, 2005.
[96] B. Liu, *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (Data-Centric Systems and Applications)*. Springer, 2008.
[97] L. B. H. Diane J. Cook, *Mining Graph Data*. Wiley-Interscience, 2006.
[98] C. Faloutsos, "Graph mining: Laws, generators and tools," in *PAKDD*, ser. Lecture Notes in Computer Science, T. Washio, E. Suzuki, K. M. Ting, and A. Inokuchi, Eds., vol. 5012. Springer, 2008, p. 1.
[99] A. Clauset, C. Moore, and M. E. J. Newman, "Structural inference of hierarchies in networks," *CoRR*, vol. abs/physics/0610051, 2006.

[100] Q. Gan and T. Suel, "Improving web spam classifiers using link structure (s)," in *AIRWeb*, ser. ACM International Conference Proceeding Series, vol. 215, 2007.

[101] C. D. M. Sepandar D. Kamvar, Taher H. Haveliwala and G. H. Golub, "Exploiting the block structure of theweb for computing," *Technical report*, 2003.

[102] R. Jin, J. Y. Chai, and L. Si, "An automatic weighting scheme for collaborative filtering," in *SIGIR*, M. Sanderson, K. Järvelin, J. Allan, and P. Bruza, Eds. ACM, 2004, pp. 337–344.

[103] J. S. Breese, D. Heckerman, and C. M. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," in *UAI*, G. F. Cooper and S. Moral, Eds. Morgan Kaufmann, 1998, pp. 43–52.

[104] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, "An algorithmic framework for performing collaborative filtering," in *SIGIR*. ACM, 1999, pp. 230–237.

[105] J. L. Herlocker, J. A. Konstan, and J. Riedl, "An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms," *Inf. Retr.*, vol. 5, no. 4, pp. 287–310, 2002.

[106] G. Linden, B. Smith, and J. York, "Industry report: Amazon.com recommendations: Item-to-item collaborative filtering," *IEEE Distributed Systems Online*, vol. 4, no. 1, 2003.

[107] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *WWW*, 2001, pp. 285–295.

[108] J. Wang, A. P. de Vries, and M. J. T. Reinders, "Unifying user-based and item-based collaborative filtering approaches by similarity fusion," in *SIGIR*, 2006, pp. 501–508.

[109] L. H. Ungar, D. P. Foster, E. Andre, S. Wars, F. S. Wars, D. S. Wars, and J. H. Whispers, "Clustering methods for collaborative filtering." AAAI Press, 1998.

[110] G.-R. Xue, C. Lin, Q. Yang, W. Xi, H.-J. Zeng, Y. Yu, and Z. Chen, "Scalable collaborative filtering using cluster-based smoothing," in *SIGIR*, R. A. Baeza-Yates, N. Ziviani, G. Marchionini, A. Moffat, and J. Tait, Eds. ACM, 2005, pp. 114–121.

[111] *Clustering for Collaborative Filtering Applications*. Press, 1999.

[112] J. F. Canny, "Collaborative filtering with privacy via factor analysis," in *SIGIR*, 2002, pp. 238–245.

[113] A. Harpale and Y. Yang, "Personalized active learning for collaborative filtering," in *SIGIR*, S.-H. Myaeng, D. W. Oard, F. Sebastiani, T.-S. Chua, and M.-K. Leong, Eds. ACM, 2008, pp. 91–98.

[114] T. Hofmann, "Latent semantic models for collaborative filtering," *ACM Trans. Inf. Syst.*, vol. 22, no. 1, pp. 89–115, 2004.

[115] L. Si and R. Jin, "Flexible mixture model for collaborative filtering," in *ICML*, T. Fawcett and N. Mishra, Eds. AAAI Press, 2003, pp. 704–711.

[116] T. Hofmann, "Collaborative filtering via gaussian probabilistic latent semantic analysis," in *SIGIR*. ACM, 2003, pp. 259–266.

[117] J. Basilico and T. Hofmann, "Unifying collaborative and content-based filtering," in *ICML*, ser. ACM International Conference Proceeding Series, C. E. Brodley, Ed., vol. 69. ACM, 2004.

[118] H. Ma, I. King, and M. R. Lyu, "Effective missing data prediction for collaborative filtering," in *Proceedings to the Special Interest Group on Information Retrieval (SIGIR2007)*, N. Kando, C. Clarke, and N. Fuhr, Eds. Amsterdam, Netherlands: ACM, 23-27 July 2007, pp. 39–46. [Online]. Available: http://www.cse.cuhk.edu.hk/∼king/PUB/SIGIR07_Ma.pdf

[119] H. Ma, H. Yang, M. R. Lyu, and I. King, "Sorec: Social recommendation using probabilistic matrix factorization," in *Proceedings to the ACM 17th Conference on Information and Knowledge Management (CIKM2008)*. Napa Valley, USA: ACM, October 26-30 2008. [Online]. Available: http://www.cse.cuhk.edu.hk/∼king/PUB/CIKM08_Ma1.pdf

[120] D. M. Pennock, E. Horvitz, S. Lawrence, and C. L. Giles, "Collaborative filtering by personality diagnosis: A hybrid memory and model-based approach," in *UAI*, C. Boutilier and M. Goldszmidt, Eds. Morgan Kaufmann, 2000, pp. 473–480.

[121] B. Mehta and T. Hofmann, "A survey of attack-resistant collaborative filtering algorithms," *IEEE Data Eng. Bull.*, vol. 31, no. 2, pp. 14–22, 2008.

[122] B. Mehta and W. Nejdl, "Attack resistant collaborative filtering," in *SIGIR*, S.-H. Myaeng, D. W. Oard, F. Sebastiani, T.-S. Chua, and M.-K. Leong, Eds. ACM, 2008, pp. 75–82.

[123] M. R. McLaughlin and J. L. Herlocker, "A collaborative filtering algorithm and evaluation metric that accurately model the user experience," in *SIGIR*, M. Sanderson, K. Järvelin, J. Allan, and P. Bruza, Eds. ACM, 2004, pp. 329–336.

[124] B. Mehta, T. Hofmann, and W. Nejdl, "Robust collaborative filtering," in *RecSys*, J. A. Konstan, J. Riedl, and B. Smyth, Eds. ACM, 2007, pp. 49–56.

[125] N. N. Liu and Q. Yang, "Eigenrank: a ranking-oriented approach to collaborative filtering," in *SIGIR*, S.-H. Myaeng, D. W. Oard, F. Sebastiani, T.-S. Chua, and M.-K. Leong, Eds. ACM, 2008, pp. 83–90.

[126] J. D. M. Rennie and N. Srebro, "Fast maximum margin matrix factorization for collaborative prediction," in *ICML*, ser. ACM International Conference Proceeding Series, L. D. Raedt and S. Wrobel, Eds., vol. 119. ACM, 2005, pp. 713–719.

[127] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM Press, 2004, pp. 168–177.

[128] J. Yi, T. Nasukawa, R. Bunescu, and W. Niblack, "Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques," in *ICDM '03: Proceedings of the Third IEEE International Conference on Data Mining*. Washington, DC, USA: IEEE Computer Society, 2003, p. 427.

[129] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," in *Proceedings COLING-04, the Conference on Computational Linguistics*, Geneva, CH, 2004. [Online]. Available: http://www.isi.edu/∼skim/Download/Papers/2004/Determining_Sentiment.pdf

[130] A. Esuli and F. Sebastiani, "Determining the semantic orientation of terms through gloss classification," in *CIKM '05: Proceedings of the 14th ACM international conference on Information and knowledge management*. New York, NY, USA: ACM, 2005, pp. 617–624.

[131] H. Kanayama and T. Nasukawa, "Fully automatic lexicon expansion for domain-oriented sentiment analysis," in *EMNLP '06*, 2006.

[132] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *EMNLP '02: Proceedings of the ACL-02 conference on Empirical methods in natural language processing*. Morristown, NJ, USA: Association for Computational Linguistics, 2002, pp. 79–86.

[133] P. D. Turney, "Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews," in *ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. Morristown, NJ, USA: Association for Computational Linguistics, 2001, pp. 417–424.

[134] L. D. Raedt and S. Wrobel, Eds., *Machine Learning, Proceedings of the Twenty-Second International Conference (ICML 2005), Bonn, Germany, August 7-11, 2005*, ser. ACM International Conference Proceeding Series, vol. 119. ACM, 2005.

[135] J. Huai, R. Chen, H.-W. Hon, Y. Liu, W.-Y. Ma, A. Tomkins, and X. Zhang, Eds., *Proceedings of the 17th International Conference on World Wide Web, WWW 2008, Beijing, China, April 21-25, 2008*. ACM, 2008.

[136] M. Sanderson, K. Järvelin, J. Allan, and P. Bruza, Eds., *SIGIR 2004: Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Sheffield, UK, July 25-29, 2004*. ACM, 2004.

[137] S.-H. Myaeng, D. W. Oard, F. Sebastiani, T.-S. Chua, and M.-K. Leong, Eds., *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2008, Singapore, July 20-24, 2008*. ACM, 2008.

[138] C. Castillo, K. Chellapilla, and D. Fetterly, Eds., *AIRWeb 2008, Fourth International Workshop on Adversarial Information Retrieval on the Web, Beijing, China, April 22, 2008*, ser. ACM International Conference Proceeding Series, 2008.

[139] *AIRWeb 2007, Third International Workshop on Adversarial Information Retrieval on the Web, co-located with the WWW conference, Banff, Canada, May 2007*, ser. ACM International Conference Proceeding Series, vol. 215, 2007.

[140] M. Najork, A. Z. Broder, and S. Chakrabarti, Eds., *Proceedings of the International Conference on Web Search and Web Data Mining, WSDM 2008, Palo Alto, California, USA, February 11-12, 2008*. ACM, 2008.

[141] A. Hinneburg, Ed., *LWA 2007: Lernen - Wissen - Adaption, Halle, September 2007, Workshop Proceedings*. Martin-Luther-University Halle-Wittenberg, 2007.