

# From Bias to Opinion: a Transfer-Learning Approach to Real-Time Sentiment Analysis

Pedro H. Calais Guerra, Adriano Veloso, Wagner Meira Jr., Virgílio Almeida  
Universidade Federal de Minas Gerais, Brazil  
{pcalais,adrianov,meira,virgilio}@dcc.ufmg.br

## ABSTRACT

Real-time interaction, which enables live discussions, has become a key feature of most Web applications. In such an environment, the ability to automatically analyze user opinions and sentiments as discussions develop is a powerful resource known as *real time sentiment analysis*. However, this task comes with several challenges, including the need to deal with highly dynamic textual content that is characterized by changes in vocabulary and its subjective meaning and the lack of labeled data needed to support supervised classifiers. In this paper, we propose a transfer learning strategy to perform real time sentiment analysis. We identify a task – *opinion holder bias prediction* – which is strongly related to the sentiment analysis task; however, in contrast to sentiment analysis, it builds accurate models since the underlying relational data follows a stationary distribution.

Instead of learning textual models to predict content polarity (i.e., the traditional sentiment analysis approach), we first measure the bias of social media users toward a topic, by solving a relational learning task over a network of users connected by endorsements (e.g., retweets in Twitter). We then analyze sentiments by *transferring* user biases to textual features. This approach works because while new terms may arise and old terms may change their meaning, user bias tends to be more consistent over time as a basic property of human behavior. Thus, we adopted user bias as the basis for building accurate classification models. We applied our model to posts collected from Twitter on two topics: the 2010 Brazilian Presidential Elections and the 2010 season of Brazilian Soccer League. Our results show that knowing the bias of only 10% of users generates an F1 accuracy level ranging from 80% to 90% in predicting user sentiment in tweets.

## Categories and Subject Descriptors

H.2.8 [Database Management]: Database Applications—*Data Mining*

## General Terms

Algorithms, Experimentation

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

KDD'11, August 21–24, 2011, San Diego, California, USA.  
Copyright 2011 ACM 978-1-4503-0813-7/11/08 ...\$10.00.

## Keywords

Social Media, Relational Learning, Sentiment Analysis, Opinion Mining, Transfer Learning

## 1. INTRODUCTION

The rise of text-based social media channels has fueled data miners with torrents of opinion-based data on the most diverse topics and entities. Blogs, microblogs and online social networks are constantly flooded with opinions about a multitude of topics such as politics, sports and other “buzz” topics that pop up daily on news media. The ability to automatically distinguish positive and negative content in streams of opinion-based data enables the creation of valuable real-time applications that monitor public opinion and summarize the aggregated sentiment of online society [13].

In this work, we argue that current state-of-the-art sentiment analysis strategies are not effective for mining opinions in this new, challenging environment. Traditionally, sentiment analysis (also known as *opinion mining*) algorithms have been designed for static and well-controlled scenarios that target analysis of reviews of products and services [26, 28]. In those scenarios, pre-defined lists of positive and negative words (i.e., lexicons) and traditional supervised machine learning techniques have been quite successful [28]. However, the real-time sentiment classification of opinion-based content on general topics discussed in social media, on a real time basis, is particularly challenging for two reasons:

1. **Data dynamicity:** Topics such as sports competitions and electoral disputes exhibit an inherent dynamic nature, caused by the *sub-events* that take place during the monitoring of the major event. Further, real time sentiment analysis needs to deal with textual data that exhibits significant *concept drift* and a *non-stationary* distribution, which both changes the characteristics of the sentiment analysis task and degrades prediction quality over time. For example, at least 50 different high-volume discussion threads arose during the US 2008 Elections [17], moreover, they could not be predicted in advance, such as the “lipstick on a pig” discussion, that resulted on the largest number of posts. Another example occurred during the 2010 Brazilian Presidential Elections when a scandal involving a close assistant of one of the candidates was unveiled in the middle of the electoral process, unleashing a large volume of unpredicted negative comments.
2. **Lack of labeled data:** Traditional machine learning strategies for sentiment analysis require training samples in order to build text-based models. However, the high volume and sparsity of sentiment stream data may render vast amounts of labeled examples unfeasible, thus compromising the potential of typical supervised learning strategies.

In this work we propose a novel strategy to address these challenges. Our strategy is based on the premises that opinion holders tend to express their opinions *multiple times*, and they are usually *consistent* in doing so. In other words, positive and negative opinions are not randomly expressed by people. For instance, someone who supports a candidate in an election will tend to post positive comments about him and negative comments about his/her adversaries on a regular basis. Technically, we say that opinion holders exhibit a varying degree of *bias*. Social theories argue that bias is an inherent feature of human behavior, which is characterized by a lack of appropriate balance, neutrality and critical doubt in argumentation [29]; one of the most clear manifestations of bias is when someone supports one side too strongly or too often [15].

In the context of social media, we develop a strategy for real-time sentiment analysis based on the *opinion holder bias prediction*, which has two main assumptions. First, users express their opinions through *endorsements*, that is, interactions among users in which one user implicitly agrees with another. We emphasize that endorsements are transactional, and such structured data is easier to deal with than textual data, enabling a more effective learning process. Second, as mentioned above, for the majority of users, bias is a *consistent, robust* characteristic of their behavior. While text content may be influenced by external factors, such as new terms that enter a topic discussion, user biases are less prone to external perturbations, unless users actually change their opinion, which is usually a relatively long process.

It is interesting to note that because opinions on a topic are not independent from the opinion holders who write them, determining user biases also addresses another objective of our study, which is to develop a *transfer learning framework* [25] strategy to analyze sentiment in social media content. Starting with the original target task  $\tau_t$  (i.e., *real time sentiment analysis*), which lacks labeled data and faces concept drift, we define *opinion holder bias prediction* as a source task  $\tau_s$  and then map solutions from  $\tau_s$  to  $\tau_t$ . It is rather intuitive that user biases may significantly improve the task of mining user opinions, but so far no sentiment analysis models have been based on that information. In addition to proposing this model, we address two key related questions: (i) How can the sociological definition of bias be implemented into a social media platform by only considering social interactions among users? (ii) How can bias information be converted into information on the sentiment that is associated with the generated content?

In summary, the main contributions of our work are as follows:

- We identify the user bias mining problem as a suitable *source task*  $\tau_s$  that is suitable for a transfer-learning strategy for real time sentiment analysis of social media content; the task focuses on the robustness and consistency of user opinions about a given topic.
- We propose and evaluate a solution to the user bias mining problem in social media in terms of a relational learning problem. Our solution exploits *endorsements* among users.
- We propose a simple unsupervised knowledge transfer approach to analyze sentiment in social media discussions based on the propagation of user bias quantifications to textual features, enabling text classifiers to use timely information on the sentiment associated with new terms arising in discussions and thus dealing with the lack of labeled training data and the non-stationary data distribution of textual content to some extent.
- We evaluate our approach using two datasets collected from Twitter [16], which is a microblogging service that disseminates user opinions and comments in real time.

We analyzed two major events in Brazil: the 2010 Brazilian Presidential Elections and the 2010 season of the Brazilian National Soccer League. Our results show that knowing the bias of 10% of users is sufficient to reach an F1 accuracy level ranging from 80% to 90% in predicting user sentiment in tweets.

## 2. RELATED WORK

In this section, we review some previous work on sentiment analysis and opinion holder bias prediction.

**Sentiment Analysis.** Sentiment analysis and opinion mining research have focused on static and well-controlled scenarios, mainly extracting opinions from product and service reviews. Two broad categories of opinion analysis strategies can be identified in the literature: lexicon-based and classification-based [11] strategies. Lexicon-based approaches use lists of words containing positive and negative terms to compute the overall polarity of the document by counting the occurrence of those terms [28]. A clear disadvantage of this strategy is that lexicons are domain-dependent and the effort needed to generate lists of words may be high.

Sentiment analysis has also been addressed as a traditional supervised learning task, with relative success. Different text classification algorithms have been applied to learn from word co-occurrences and linguistic features to determine the sentiment contained in documents [14, 28]. Moreover, it has been used conjunction with pre-defined lexicons to assess sentiment in political and movie review blogs [22].

More recently, the increasing availability of opinion-based data in real time has motivated some studies that have analyzed sentiments in streaming data, specially over Twitter. Some approaches are as simple as the manual classification of tweets and lexicons of positive and negative words to monitor the debate performance of candidates in the 2008 US Elections [6] [24]. While lexicons may provide sentiment analysis on an aggregated level, their coverage in terms of content is usually low because in complex contexts such as elections and sports, content is often ironic, contains subtle comments and refers to specific terms that only make sense at a specific time; it often lacks expressions of clear polarity such as “I love it” or “I hate it”.

Standard classification techniques have also been tested on Twitter [2], as have stream classification techniques such as the Multinomial Naive Bayes and the Stochastic Gradient Descent [3]. The major drawback of these approaches is that they require labeled data, which are very costly to obtain on a regular basis in a volume large enough to properly address concept drift. Further, active and semi-supervised strategies aim at reducing labeling and training efforts [5], but they still require training data to be sampled from a stationary distribution. In addition to that, in microblogs such as Twitter, the small document lengths restrict the possibility of using co-occurrence among terms and other standard text mining techniques to assign classes from an initial set of labeled documents.

**Bias.** Bias has been extensively studied by sociologists as a characteristic of people or entities that exhibit one or more of the following characteristics [29]: (i) a lack of proper balance and neutrality in argumentation, (ii) a lack of proper critical doubt, (iii) a particular position of the arguer regarding a subject, and (iv) a personal interest from the arguer in the outcome of the argument or discussion. We can find bias in almost any scenario where opinions are expressed, and one of the most common types of bias is political or ideological bias [15]. In news media, selection bias guides decisions regarding of events are covered, while description bias is related to the veracity of the media coverage [7]. Given the subjective nature of bias, one key question here involves measuring

bias. One strategy to measure media bias is to count the number of times a particular media outlet cites various sources, and compare this to the citation rates of those sources by congressmen; this approach, for instance, unveiled a strong liberal bias in the US news media [23]. In social media, bias has mainly been studied with respect to blogs. Some studies categorize political blogs according to their political bias and analyze the communication patterns and network structure that different political views induce [1]. Bias in political blogs has also been used to predict the bias of political articles in the online news media [10], by counting the number of liberal and conservative blogs that cite each article. These studies focus on the study of bias on the domain of blogs and on two main groups of blogs (liberal and conservative). In this paper, we extend these studies by analyzing bias on a microblog platform and considering an additional domain (i.e., sports).

**Our Work.** This paper lies between the two fields described above. We propose a *transfer learning* approach for real time sentiment analysis that connects both the abovementioned tasks. Transfer learning is applicable when a task is hard to solve (e.g., due to lack of labeled data or when the data becomes quickly outdated), but the task may benefit from knowledge obtained from similar tasks or domains [18, 25]. Experiments and techniques that apply transfer learning to sentiment analysis are relatively new; some recent works developed *domain adaptation* techniques, which comprise a special case of transfer learning when labeled data is available for one domain but absent in the target domain. This technique has been applied to transfer knowledge acquired from the analysis of opinions about a certain class of products (e.g., movie reviews) to classify documents about other products (e.g., books) [4, 18].

Transfer learning is an appealing strategy to provide real time sentiment analysis, because the availability of labeled data is even more limited than in classical scenarios such as product review. Moreover, text-based models can become quickly outdated because of the dynamic nature of online discussions. Furthermore, a similar domain with labeled data is not easy to find, as any comparable domain would present the same drawbacks. We then present a novel transfer learning approach for real time sentiment analysis: instead of learning text-based models to predict document polarity, we solve a different (but related) problem (i.e., social media user bias prediction) and then *transfer* the knowledge acquired to sentiment analysis algorithms. In contrast to recent domain adaptation efforts, in which the source task still involves dealing with textual data, the source task is chosen to provide consistent, robust patterns of user behavior in order to support the target task.

### 3. QUANTIFYING USER BIAS IN SOCIAL MEDIA

In this section we describe how we use social media endorsements to quantify user bias towards a topic. We start by discussing which types of information may indicate endorsement, and then we present our model of user bias prediction, we include an the opinion agreement graph and discuss how it can be used to quantify user bias.

#### 3.1 User Endorsements in Social Media

As mentioned above, user endorsements express a user’s support of a specific opinion which coincides with that user’s bias. One piece of basic information that is frequently exploited as evidence of potential agreement between social network users is the social relationship that connects them. However, these relationships may not represent endorsements *per se*, because bias is topic-dependent, and two persons who are connected through a relationship may

agree regarding a certain topic but disagree with respect to another topic. Even within a specific topic, social ties are not definitive evidence of agreement between users; in the 2010 Brazilian Presidential Elections, for example, about half of the users who followed one of the top two candidates in Twitter also followed the other. While this may indicate neutrality in some cases, we believe that a significant fraction of such users where actually following his or her preferred candidate and monitoring the other.

As compared to social ties, *social interactions* (such as the exchange of messages between users) have been shown to be stronger indicators of link strength among users [30]. Furthermore, interactions are usually associated with *content*; thus, they may be categorized into specific topics, unlike social ties.

Most social media platforms support *endorsement interactions*, through which a user explicitly agrees with other user with respect to certain content. In Twitter, *retweets* are endorsements in which a user propagates a message posted by another user to their list followers, while in Facebook users may explicitly support a content (and, indirectly, the user who generated it) by clicking on the “Like” button. Endorsements may be represented as a directed graph, where an edge  $(u,v)$  represents that user  $u$  has endorsed user  $v$ . Figure 1 shows an example of an endorsement graph. Nodes A and B are highlighted; their significance will be clarified later in this section.

#### 3.2 Modeling User Bias Prediction

Our proposal to model user bias is based on the premise that similar users share a similar bias. Thus, before we analyze sentiments, we first determine the most similar users based on individual endorsements.

We model the user bias prediction problem as a *relational learning problem* [12]. More specifically, we treat it as a *within-network classification problem*, where the goal is to classify the nodes of a partially labeled network [21]. Assuming that there are  $K$  possible sides that a user may support regarding a given topic (e.g., candidates, parties or football clubs), we represent each user  $u$  as a *bias vector*  $\vec{B}_u = [B_{u1}, \dots, B_{uK}]$ , where each  $B_{ui}$  quantifies the bias of user  $u$  towards side  $i$ . We expect that neutral users are equally distant (or close) to all sides, while biased users should be closer to one side (or a selection of sides) than others.

Formally, given  $K$  sides, a set of users  $U$ , a set of relationships  $E$ , and a set  $A \subset U$ , which contains users with known bias, we define the problem of predicting user bias as the estimation of  $B_{ui}, \forall u \in U, 1 \leq i \leq K$ .

Most methods proposed so far for addressing this problem infer the missing information in aggregate based on the hypothesis that linked or nearby nodes are likely to have the same labels [20]. However, a major challenge in our scenario is to determine the training set  $A$ , not only because bias quantifications are usually not available as ground truth, but also because quantifying bias is a subjective assessment for humans. More specifically, judging whether social media users are “too biased” towards one side of the discussion or “just a bit biased” is simply not feasible. That is, manually assigning a bias measure for a significant set of users is costly and not practical.

#### 3.3 The Opinion Agreement Graph

We address the challenge of determining user bias by introducing a new graph, which we call the Opinion Agreement Graph (OAG), where neighbors (i.e., users that are close in the graph) tend to be similar. The rationale behind OAG is that the weight of an edge between two users should quantify both the intensity of their endorsement relative toward a given set of users, and the intensity of the en-

dorsement of a given set of users toward them. We call those measures active and passive similarity, respectively, and they are calculated using frequent pattern mining techniques [27], as follows. The active similarity comes from a “database”  $\mathcal{D}_\alpha$  of “transactions” where each transaction contains the users who endorsed a given user. For each pair of users  $(u, v)$  we define its active similarity  $\alpha(u, v)$  as the *lift* of pair  $(u, v)$  in  $\mathcal{D}_\alpha$ . Lift is a standard metric of interest in association rule mining that captures how surprising a relationship among two items is, by comparing the frequency of their co-occurrence to the *expected* frequency of their co-occurrence, assuming that they are independent from one another [27]. Notice that the use of lift is compatible with the sociological definition of bias, which indicates biased behavior when someone supports one side or someone too strongly or too often [15].

The passive similarity  $(\rho(u, v))$  of each pair of users  $(u, v)$  is calculated in the same way; the only difference is that  $\mathcal{D}_\rho$ , which is the passive database, contains the users endorsed by a given user in each transaction. Passive similarity is important for measuring bias using data beyond’s user direct endorsements the direct control of the user: even if he/she does not endorse any content/user, we can assess his/her bias using the bias of users who endorse him/her. This is particularly useful to quantify the bias of news media profiles, which usually never endorse any content but must be evaluated based on how they are interpreted by the other users.

We then define the *Opinion Agreement Graph*  $G(V, E)$  which contains pair-wise similarity information, where  $V$  is the set of vertices so that each vertex is associated with a user (item) from  $\mathcal{D}_\alpha \cup \mathcal{D}_\rho$ , and  $E$  is the set of weighted edges that represent relationships between pairs of users, with weight  $E_{u,v} = \text{mean}(\alpha(u, v) + \rho(u, v))$ .

In Figure 2, we show the resulting OAG calculated from endorsements  $(u, v)$  from Figure 1. Solid lines connect users that share an active bias similarity, and dashed lines represent passive biases.

Note that the semantics of an edge in the OAG is richer than a directed edge in the original endorsement graph: instead of simply showing that a given user endorsed another user, an edge in the OAG captures the strength of the similarity relationship of two users based on the set of users *as a whole*.

It is interesting to note that the different criteria used to define edges generate a graph topology that makes explicit relationships that were hidden in the original endorsement graph. Node 6, which is apparently a sink in Figure 1, now emerges as an important bridge between nodes A and B. Nodes 1, 2, 3 and 4, which share only one edge in the endorsement graph (i.e., an endorsement of user 2 by user 1), now form a dense sub-graph because they commonly endorse nodes 5 and A. The OAG also connects nodes that are not mutually reachable in the endorsement graph, such as pairs (5,6), (8,10), and (6,11). Node 7, however, does not appear in OAG, because it does not have common neighbors in terms of either incoming or outgoing edges.

**Attractors.** There usually are a few users that are clearly biased toward one or more sides of a discussion, based on prior knowledge. For example, in a political discussion, the official profiles of candidates and parties are expected to only express opinions that are favorable to their side. We label users whose bias is clearly identifiable as representative of a particular side in a discussion as *attractors*. In Figures 1 and 2, nodes A and B are assumed to be attractors. We associate each side  $i$  of a topic discussion with a set of attractors  $A_i = [A_{i1}, \dots, A_{ij}]$ , which are the nodes associated with the users chosen to represent that side.

Attractors serve as reliable sources of bias knowledge, and they allow us to determine the bias of other users. The bias of each node is its *proximity* from attractors that represent that side to all users

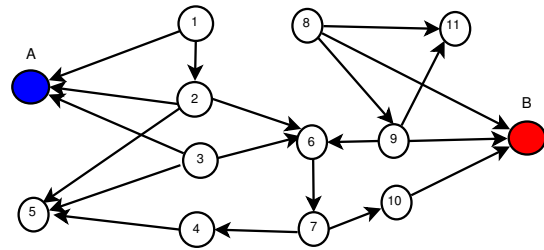


Figure 1: An hypothetical endorsement graph.

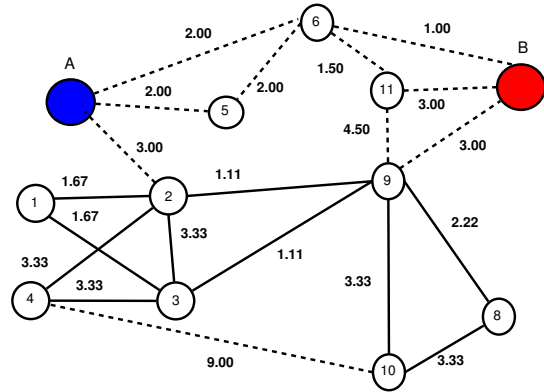


Figure 2: An OAG for endorsements from Figure 1, which connects users who endorse and are endorsed by common sets of users, respectively represented by solid and dashed lines. Edge weights are calculated based on the lift of the relationship between  $u$  and  $v$ . For example, nodes 9 and 11 are endorsed by the same users at a rate 4.5 times higher than expected. In contrast to the endorsement graph, edges in AOG are undirected and capture the global opinion of the whole network on the proximity of the users they connect.

in  $U$ . There are several strategies to determine this proximity; we used random walk to measure of proximity among nodes, due to its capacity to capture the multi-faceted relationship between nodes in a graph [8]. Formally,

$$B_{ui} = \text{RandomWalk}(G, A_i, u) \quad (1)$$

Our approach is in accordance with recent studies that demonstrate that using few seeds can be an effective strategy for propagating labels in graphs [9, 19]. Our scenario is particularly suitable for such a strategy because of the high consistency of the relationships we create, which indicate that the connected pair of users commonly endorse and/or are commonly endorsed by a *group* of users, in contrast to the original endorsement graph, in which any user endorsing another user creates a path in the graph. Furthermore, the edges in OAG are undirected, what turns the analysis of connectivity among users simpler and more effective.

**Neutral user normalization.** In a traditional within-network learning setting, a node linked to a large number of neighbors from class  $C_1$  and a small number of neighbors from class  $C_2$  will mostly likely be assigned to class  $C_1$ , which will be acceptable for most applications. In our scenario, however, we must consider that different sides of a discussion may have different strength levels in a social media platform, and thus they may differentially affect user bias values. In other words, it is not reasonable to directly compare proximities calculated from different starting distributions; the network, as a whole, may be more inclined to a subset of sides. To address this problem, we define the *neutral user*  $\vec{N}$  as a user who

endorses all other users, and has been endorsed by all other users. The bias vector  $\vec{N}$  captures the global tendency of the network and is used to adjust all biases by a component-wise division:

$$\vec{B}_u = \frac{\vec{B}_u}{\vec{N}} \quad (2)$$

In the OAG shown in Figure 2, let A and B be attractors that represent two sides of a discussion, e.g., two candidate profiles. To predict the bias of users 1 to 11 regarding those two sides, we compute the random walk proximity of A and B towards each of those users. The resulting bias vector of each user is plotted in Figure 3. Note that nodes 9 and 11 are closer to B, while node 2 has a clear preference for attractor A. In particular, the direction of the bias vector indicates the bias of the user with respect to each attractor, whereas its magnitude captures the intensity of user activity. Thus, we have different patterns of neutrality: users who strongly support two or more sides, and users who strongly advocate against all sides.

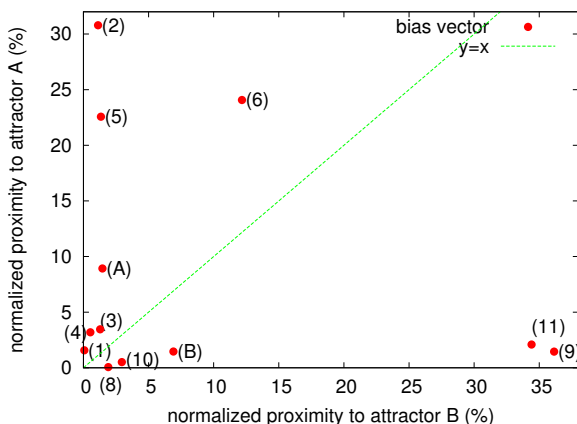


Figure 3: Bias vectors calculated as random walks from attractors.

#### 4. CASE STUDY: POLITICAL AND SPORTS DISCUSSIONS IN TWITTER

We use the strategy introduced in Section 3 to quantify user bias in online discussions about two major events that took place in Brazil in 2010: the Presidential Elections and the Brazilian First Division Soccer League season. Both events were heavily debated on *Twitter*, which is a popular microblogging system that allow users to exchange short messages that are instantly made available to the set of users who follow the message’s originator [16]. It is worth noticing that our proposal assumes that there is some polarity in the discussions as a basis for determining user bias, and these two scenarios are clearly polarized, discussed below.

The Brazilian presidential election campaign was held from June to October 2010, and two candidates led the polls since July 2010: Dilma Rousseff and Jose Serra. Rousseff and Serra each had more than 500,000 followers on Twitter, and both tried to use the system as one of the main means of communication with their voters. After the election came to a second round, Dilma Rousseff won the runoff with 56% of the vote.

In Table 1, we show a general overview of the two datasets considered in this work. The data collection of each dataset was performed using the Twitter API<sup>1</sup>; data was collected by focusing on

<sup>1</sup>available at <http://apiwiki.twitter.com/>

the entities involved in each topic. In the `Elections-BR` dataset, the entities were the names of the two candidates starting in July 2010 and leading up to the 2010 Brazilian Presidential Elections (i.e., Dilma Rousseff and Jose Serra). To build the `Soccer` dataset, we considered the 12 most popular Brazilian soccer teams, which played on the Brazilian 2010 First Division Soccer League. In both domains, we can observe that a significant fraction of users endorsed tweets of others, using the retweet mechanism that propagates a tweet from one user to all of another user’s followers.

Table 2 shows some statistics from the Opinion Agreement Graph for both datasets. The graphs considered only users who were involved in at least five endorsements. Note that a fraction of edges are a combination of active and passive similarities.

	<code>Elections-BR</code>	<code>Soccer</code>
number of nodes (users)	34,678	172,398
number of active bias edges	460,808	4,426,235
number of passive bias edges	607,331	886,895
number of active+passive bias edges	143,052	146,935
average degree	21.41	35.72
average edge weight	161.27	306.33

Table 2: Opinion Agreement Graph Statistics.

In their respective domains, the official profiles of candidates and soccer clubs are natural attractors. After computing the vector of proximities relative to these attractors and plotting the results in Figures 4 and 5, the high degree of polarization in the discussions on both topics becomes clear. Regarding the Brazilian Elections (Figure 4), at least 76% of users have a clear inclination towards one of the candidates. In the case of Twitter users who comment about soccer, the vast majority of users have a clear inclination towards one specific team (Figure 5).

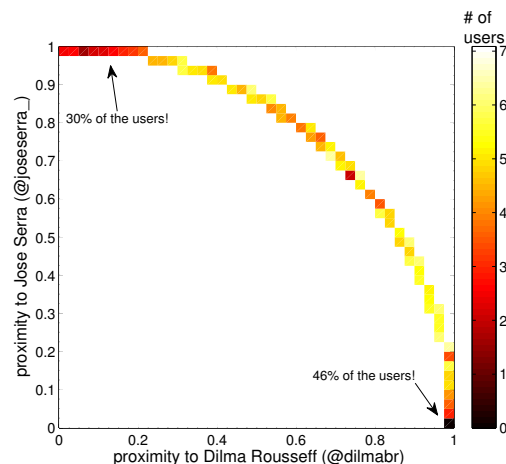


Figure 4: Bias vectors for the 2010 Brazilian Presidential Elections discussions. At least 76% of users are much closer to one side than the other, indicating a clear political bias.

**Bias consistency.** We empirically evaluate the robustness and consistency of our bias quantification of user tendencies. The intuition is that, given that users exhibit consistently a biased behavior, we should infer similar bias of a user based on two different samples of his/her endorsements.

We divided the set of endorsements into two sets; the first set comprised all of the endorsements observed during the first half of the data collection period, and the second set comprised the other

	Elections-BR	Soccer
period	2009-12-18 to 2010-10-03	2010-05-08 to 2010-09-07
entities	2 candidates	12 football clubs
number of tweets	7,707,192	8,828,520
number of retweets	2,511,779 (32.6% of tweets)	1,866,593 (21.1% of tweets)
number of users	1,022,396	1,633,537
number of users retweeting posts	489,214 (47.8% of users)	584,685 (35.7% of users)
number of retweeted users	179,441 (17.5% of users)	337,352 (20.6% of users)
avg. number of unique users retweeted by each user	16.41	5.54
avg. number of unique users that have retweeted each user	6.24	3.43

Table 1: General Overview of the Datasets.

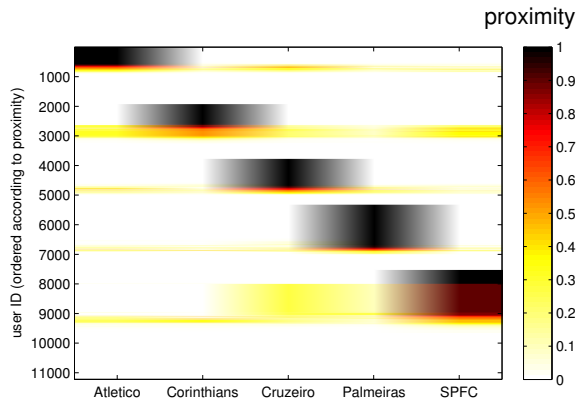


Figure 5: Bias vectors of users towards soccer teams. Supporters of Brazilian teams Atletico, Corinthians, Cruzeiro, Palmeiras and SPFC are identified. Users close to all teams are likely to be commentators or neutral media vehicles.

half. We then calculated the bias vector for all users in each set. We computed two bias measures for the second set. One considers the original pairs  $(u, v)$  of endorsements, and the other considers randomized endorsements. We then compute the cosine distance between each pair of vectors associated with each user from both sets, for users who were involved in endorsements in both sets. In Figure 6, we plot the cumulative distribution of the cosine distance for the two scenarios (that is, the original and randomized sets). Due to space constraints, we only show our analysis for the political discussion, but results were similar for the sports discussions.

Note that for more than 90% of users, vectors calculated from the two samples presented a high cosine similarity, greater than 0.80. However, when we compare bias vectors generated from the first set to the randomized samples from the second set, they are no longer similar. This means that, for the majority of users, the direction of their bias vectors does not change significantly across time but rather reflects the robustness and consistency of user behavior. Thus, it is not crucial to update the Opinion Agreement Graph and bias quantifications as new endorsements are observed, and as such, this task may simply be done at regular intervals. In addition, if the bias of a small sample of users is known, sentiment analysis based on that information is already possible, as we will show in the next section.

## 5. EXPLOITING USER BIAS FOR REAL-TIME SENTIMENT ANALYSIS

In this section we introduce our transfer-learning strategy for analyzing sentiments. The input is a constant flow of triplets in the

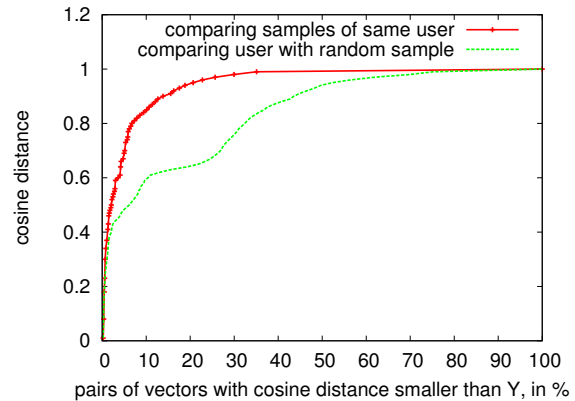


Figure 6: Small changes in user bias vectors calculated from different samples indicate robustness of user bias patterns in the ELECTIONS-BR dataset.

form of  $\langle author, d, e \rangle$ , where *author* is the opinion holder who wrote message *d*, which mentions an entity *e*. The goal is to predict the *polarity*  $p = \{+, -\}$  of message *d* toward *e*. As discussed, there are several challenges in performing such a task and the dominant approach relies on extracting textual patterns from messages and exploiting these patterns to predict polarity.

User bias, in contrast, provides information that is more robust to these challenges, as it is more consistent than typical textual information (recall Figure 6 from Section 4). Thus, we propose an alternate approach that is based on transfer learning [25]. More specifically, we first solve a *source task* (denoted as  $\tau_s$ ), which involves estimating the bias of some users. This bias information is then *transferred* to the *target task* (denoted as  $\tau_t$ ), which is to predict the polarity of messages referencing entity *e*.

**Propagating bias across terms.** We transfer information from task  $\tau_s$  to task  $\tau_t$  by assuming that term *t* will be *positive* toward entity *e* if it is adopted more frequently by users biased toward entity *e* than by users of different sides in tweets that mention *e*. Similarly, *t* will be *negative* to entity *e* if it is adopted by a large number of users who oppose that entity, in contrast with the number of supporters of *e*. Neutral content is expected to be endorsed by both sides. To validate this intuition, in Figure 7 we plot the bias vector associated with users that referred to three different web pages in their tweets: a YouTube video with positive comments about Jose Serra (Figure 7a), an official video from Dilma Rousseff’s campaign (Figure 7b), and a general news article about the 2010 Presidential Elections (Figure 7c).

In order to transform user bias into term bias, we take into account the bias vector associated with each user that used term *t*. A

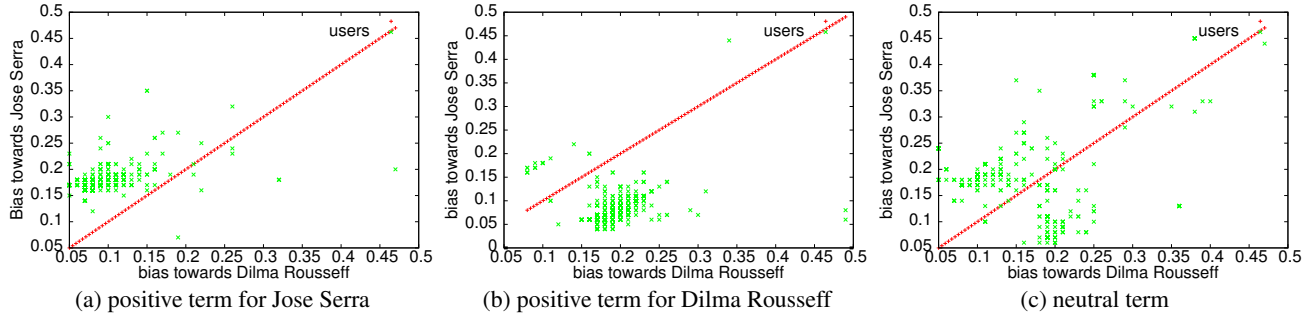


Figure 7: User bias vectors for three different contents, which show that user bias is a good predictor of term polarity.

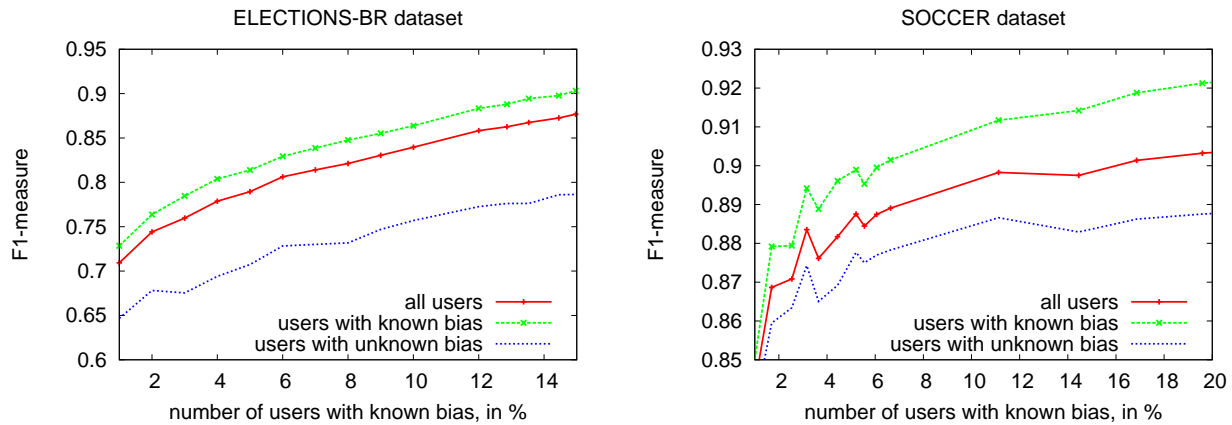


Figure 8: F1-accuracy level for different ratios of users with known bias. Left - ELECTIONS-BR dataset. Right - SOCCER dataset. As the ratio of users of known bias increases, the F1-measure increases, even for tweets posted by users with unknown bias.

possible unsupervised approach is to compute the sum vector of all users present in the Opinion Agreement Graph that refer to entity  $e$  by adopting term  $t$ :

$$\vec{B}_{t,e} = \sum_{u \in V} \vec{B}_u \quad (3)$$

Note that this computation *propagates* user bias information to all messages that contain at least one term adopted by a user with known bias, thereby revealing the judgement of the content produced by users with unknown bias. This is important because it is expected that information on user bias will be available for only a portion of users, since many users are never involved in endorsement interactions, as shown in Table 1.

**Dealing with concept drift.** When a term is adopted for the first time, it will have the same bias as the corresponding user who adopted it. As new messages pass through the stream,  $\vec{B}_{t,e}$  is updated incrementally. As such, users collectively judge new terms, referring to them (or not) in their messages. To predict the polarity of a message  $d$ , we first convert the bias vector of each term present in a message into *polarity probabilities*. Given the bias vector  $\vec{B}_{t,e}$ , and that  $\vec{B}_{t,e,e}$  represents the strength of component  $e$  in  $\vec{B}_{t,e}$ , we calculate the probability that term  $t$  refers positively to entity  $e$  according to Equation 4.

$$\hat{p}(\text{polarity} = +|t, e) = \frac{B_{t,e,e}}{\|\vec{B}_t\|} \quad (4)$$

Note that we compare the strength of bias to  $e$  in  $\vec{B}_{t,e}$  with the magnitude of the bias vector. Specifically,  $\hat{p}(\text{polarity} = -|t, e)$  may be calculated as  $1 - \hat{p}(\text{polarity} = +|t)$ . To predict message polarity, we may adopt various strategies to combine those probabilities. Limited to 140 characters, Twitter messages are short, thus, we exploit a simple strategy for predicting message polarity, which is to consider the term of highest polarity in each tweet:  $\text{polarity} = \text{argmax}(\hat{p}(\text{polarity} = x|t))$ .

In Figure 8, we analyze the performance of our transfer learning approach as the fraction of users whose known bias varies. We report performance numbers using the F1 measure. To generate ground truth with respect to messages, we combined manual labeling with automatic labeling for messages containing tags that clearly indicated a preference for a specific entity. To make our evaluations fair, we removed all tags used to generate our labels from message content. We can see that the F1-measure increases as the proportion of users with known bias increases, up to a point at which F1 stabilizes. When the bias of 15% of users commenting on politics is known, the F1-measure equals 85%, while in the corresponding case for soccer, F1 is 90%. Note that the F1-measure for posts from users with unknown bias also increases as we transfer bias from a greater number of users, what further demonstrates the applicability of our user-term bias transfer approach.

**Comparison with SVM.** We now compare the F1-measure provided by our bias-based sentiment analysis model against the same metric provided by a typical SVM classifier. We chose SVM because it has already been successfully applied to various sentiment

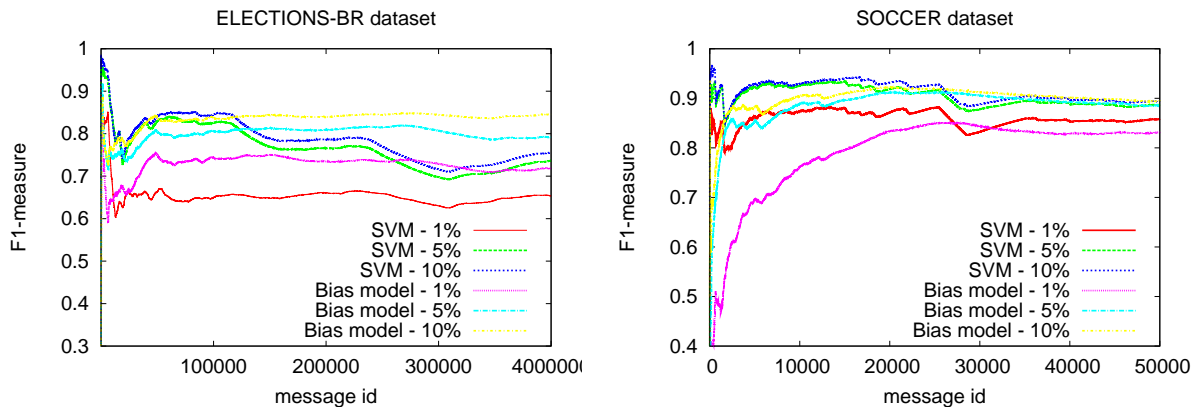


Figure 9: Bias-based model versus SVM classifier. Left - ELECTIONS-BR dataset. Right - SOCCER dataset.

analysis application scenarios, including the analysis of tweets [2, 28]. In order to execute this experimental comparison, we split each dataset into two partitions. The first partition is used for training, and it is comprised of the first 10% of tweets from each dataset. The second partition is used to validate each approach. Our comparison involves the execution of different training configurations. More specifically, each execution uses 10%, 50% or 100% of the training partition. For SVM, the training partition was used to train textual-based models, while for our bias-based model, we only considered endorsements in order to compute the OAG and generate bias assessments for users. When we compared the results on a chronologically ordered set of labeled tweets (i.e., the test partition), as shown in Figure 9, some important observations arise. We can note that the SVM F1-measure decreases across time for both datasets, which is evidence of changes in the textual feature distribution. In contrast, the bias-based sentiment classifier is able to maintain a stable F1-measure, as it incrementally incorporates bias information on new terms by propagating user bias. Indeed, the bias model performs better than SVM for the ELECTIONS-BR dataset, and it exhibits a similar F1-measure for the SOCCER dataset, even though it does not require labeled textual data, as in SVM.

**Sentiment during a soccer match game.** In order to demonstrate the capability of our sentiment analysis approach in analyzing the reactions of microblog users during live events, we have chosen to follow the buzz generated on Twitter during one of the most exciting matches of the Brazilian 2010 Soccer Season, which pitted team “Cruzeiro” against “Atletico Mineiro”, two fierce rivals. Atletico won by 4-3, scoring three goals in the first half. In Figure 10, we plot positive and negative comments for each team during the game. Positive and negative comments were shown along the Y-axis. We have highlighted the timing of the goals with numbers (in black for Atletico Mineiro and in white for Cruzeiro). Note that positive comments tend to coincide with the timing of the goals, and are targeted at the team who scored. At 36’, Cruzeiro conceded a third goal, and at that point, we detected a burst of negative comments about the team. Note that due to Atletico Mineiro’s good game, almost no negative comments toward the team were observed during entire game. After the game finished, a spike of positive comments about Atletico was detected, at 95’.

## 6. CONCLUSIONS AND OUTLOOK

Real time sentiment analysis is a difficult task; labeled data is usually not available to support supervised classifiers, and debate about monitored topics may turn into unpredictable discussions.

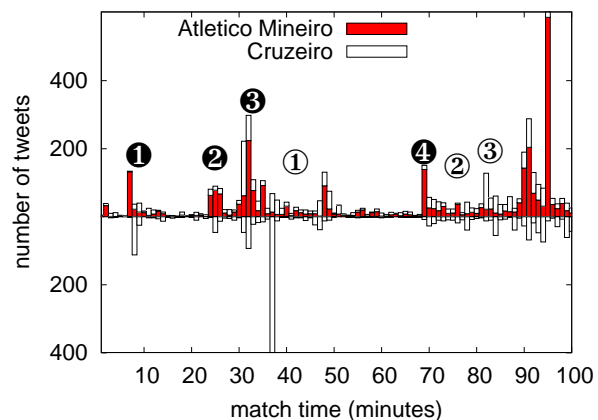


Figure 10: Stacked histogram showing sentiment variation during the Brazilian National Soccer League game Atletico Mineiro 4 – 3 Cruzeiro. Comment peaks coincide with goals.

Our contribution to the field of sentiment analysis addresses those challenges by proposing a novel transfer learning approach to topic-based real time sentiment analysis. We identify a suitable source task (that is, *opinion holder bias prediction*), which is motivated by sociological theories that argue that humans tend to have consistently biased opinions. We model this problem as a relational learning task by leveraging the network structure induced by the mutual endorsements among social media users. We then devise a simple and effective knowledge transfer strategy that propagates user bias to terms associated with user content. Then, term biases are combined to compute the overall content polarity.

Our work has demonstrated that the consistency in human bias enables robust, real time sentiment analysis without labeled text data, on topics in which polarization among user opinions occurs (such as politics and sports). By knowing the bias of only 10% of users who commented about those topics in Twitter, we were able to correctly classify the polarity of 80% to 90% of the tweets.

Our next step is to incorporate bias information into state-of-the-art sentiment analysis algorithms. Furthermore, we want to evaluate the applicability of our strategy to contexts with lower degrees of polarization, such as sentiment analysis of product review data, and we hope to derive theoretical guarantees regarding performance based on the degree of polarization among opinion holders.

## Acknowledgments

This work was partially supported by CNPq, CAPES, FAPEMIG, FINEP, InWeb and by UOL ([www.uol.com.br](http://www.uol.com.br)) through its Research Scholarship Program, Proc. Number 20110215235100.

## 7. REFERENCES

- [1] L. A. Adamic and N. Glance. The political blogosphere and the 2004 U.S. election: divided they blog. In *LinkKDD '05: Proceedings of the 3rd international workshop on Link discovery*, pages 36–43, New York, NY, USA, 2005. ACM.
- [2] A. Bermingham and A. F. Smeaton. Classifying sentiment in microblogs: is brevity an advantage? In *Proceedings of the 19th ACM international conference on Information and knowledge management, CIKM '10*, pages 1833–1836, New York, NY, USA, 2010. ACM.
- [3] A. Bermingham and A. F. Smeaton. Crowdsourced real-world sensing: sentiment analysis and the real-time web. In *AICS 2010 - Sentiment Analysis Workshop at Artificial Intelligence and Cognitive Science*, 2010.
- [4] J. Blitzer, M. Dredze, and F. Pereira. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 440–447, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [5] O. Chapelle, B. Schölkopf, and A. Zien. *Semi-supervised learning*. MIT Press, 2006.
- [6] N. A. Diakopoulos and D. A. Shamma. Characterizing debate performance via aggregated twitter sentiment. In *Proceedings of the 28th international conference on Human factors in computing systems, CHI '10*, pages 1195–1198, New York, NY, USA, 2010. ACM.
- [7] J. Earl, A. Martin, J. D. McCarthy, and S. A. Soule. The use of newspaper data in the study of collective action. volume 30, pages 65–80. *Annual Review of Sociology*, 2004.
- [8] F. Fouss, A. Pirotte, J.-M. Renders, and M. Saerens. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. *IEEE Trans. on Knowl. and Data Eng.*, 19:355–369, 2007.
- [9] B. Gallagher, H. Tong, T. Eliassi-Rad, and C. Faloutsos. Using ghost edges for classification in sparsely labeled networks. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '08*, New York, NY, USA, 2008. ACM.
- [10] M. Gamon, S. Basu, D. Belenko, D. Fisher, M. Hurst, and A. C. König. Blews: Using blogs to provide context for news articles. In *In Proceedings of the 2nd International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2008.
- [11] S. Gerani, M. J. Carman, and F. Crestani. Proximity-based opinion retrieval. In *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval, SIGIR'10*, pages 403–410, New York, NY, USA, 2010. ACM.
- [12] L. Getoor and B. Taskar. *Introduction to Statistical Relational Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2007.
- [13] B. J. Jansen, M. Zhang, K. Sobel, and A. Chowdury. Twitter power: Tweets as electronic word of mouth. *J. Am. Soc. Inf. Sci. Technol.*, 60:2169–2188, November 2009.
- [14] W. Jin, H. H. Ho, and R. K. Srihari. Opinionminer: a novel machine learning system for web opinion mining and extraction. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '09*, New York, NY, USA, 2009. ACM.
- [15] M. Kienpointner and W. Kindt. On the problem of bias in political argumentation : an investigation into discussions about political asylum in Germany and Austria. *Journal of Pragmatics*, 5(27):555–585, 1997.
- [16] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *WWW '10: Proceedings of the 19th international conference on World wide web*, pages 591–600, New York, NY, USA, 2010. ACM.
- [17] J. Leskovec, L. Backstrom, and J. Kleinberg. Meme-tracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '09*, pages 497–506, New York, NY, USA, 2009. ACM.
- [18] T. Li, V. Sindhwani, C. H. Q. Ding, and Y. Z. 0005. Bridging domains with words: Opinion analysis with matrix tri-factorizations. In *SDM*, pages 293–302, 2010.
- [19] F. Lin and W. W. Cohen. Semi-supervised classification of network data using very few labels. In *ASONAM*, pages 192–199, 2010.
- [20] S. A. Macskassy and F. Provost. A simple relational classifier. In *Proceedings of the Second Workshop on Multi-Relational Data Mining (MRDM-2003) at KDD-2003*, pages 64–76, 2003.
- [21] S. A. Macskassy and F. Provost. Classification in networked data: A toolkit and a univariate case study. *J. Mach. Learn. Res.*, 8:935–983, December 2007.
- [22] P. Melville, W. Gryc, and R. D. Lawrence. Sentiment analysis of blogs by combining lexical knowledge with text classification. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '09*, New York, NY, USA, 2009. ACM.
- [23] J. Milyo and T. Groseclose. A measure of media bias. Working Papers 0501, Department of Economics, University of Missouri, Jan. 2005.
- [24] B. O'Connor, R. Balasubramanyan, B. Routledge, and N. Smith. From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the Int'l AAAI Conference on Weblogs and Social Media*, 2010.
- [25] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22:1345–1359, 2010.
- [26] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135, 2008.
- [27] P. Tan, Steinbach, M., and V. Kumar. *Introduction to Data Mining, (First Edition)*. Addison-Wesley Longman Publishing Co., 2005.
- [28] P. D. Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting on Assoc. for Computational Linguistics, ACL '02*, Morristown, NJ, USA, 2002. Assoc. for Computational Linguistics.
- [29] D. Walton. Bias, critical doubt, and fallacies. Number 28, pages 1–22. *Argumentation and Advocacy*, 1991.
- [30] R. Xiang, J. Neville, and M. Rogati. Modeling relationship strength in online social networks. In *Proceedings of the 19th international conference on World Wide Web, WWW '10*, pages 981–990, New York, NY, USA, 2010. ACM.