Modeling, Quantifying and Visualizing Media Bias on Twitter

ANAM ZAHID, (Member, IEEE), MAHAM NASIR KHAN, AHMER LATIF KHAN, FAISAL KAMIRAN, AND BILAL NASIR

Department of Computer Science, Information Technology University of the Punjab, Lahore 54600, Pakistan
Corresponding author: Anam Zahid (anam.zahid@itu.edu.pk)

ABSTRACT News media garner a lot of attention regarding the subjectivity of their reporting. News media bias is of immense interest to various individuals, as the systematic preference of an entity can invoke its support and public actions. These inclinations, although apparent, hinder the true facts. The identification and quantification of media bias is one of the most important metrics in reference to bias assessment in media and general public. In this paper, we present a principled approach to quantify media bias along with insightful visualizations for popular media sources using their tweets. We use the concept of a mini-world of \( N \times M \) matrix to model the sources and entities of interest, where the tweet counts and respective polarities over a specified time period are the values. Direct comparisons between these two are not as meaningful due to the neglect of inherent characteristics of sources and entities. Thus, we define coverage and statement scores as properly normalized measures of tweet counts and polarity rates. Furthermore, we present a statistically consistent model of neutral tweet counts and polarity rates, using which we define the absolute coverage and statement bias of each source-entity pair. We illustrate our approach on two data sets capturing tweets on 1) Prime minister candidates of top political parties of Pakistan in the 2018 general election 2) Paris and Beirut bombings in 2015 by different news sources. The results indicate that our model is generalizable i.e. it can be applied to different entities/sources and in consistent with previous studies.

INDEX TERMS Bias analysis, information retrieval, media bias, social media.

I. INTRODUCTION

The term “bias” is defined as an inclination towards a subject or an entity without any definite facts or proofs. A news media source is said to be biased when they systematically incline towards certain events or stories and present them in a way (e.g., positive or negative) [1] that tends to systematically shape the opinion of the people to favour one side of the political spectrum over the other. These inclinations towards an entity by the news media are governed mainly by financial and political ties along with some additional factors [2]. Media sources can also exhibit biases with respect to socially sensitive characteristics like race, religion, age, and gender [3]. Media bias is identified through three different aspects [4] namely

1) Gate-keeping or Selection Bias: The measure of media bias where it is decided whether a story is to be published or not.

The associate editor coordinating the review of this manuscript and approving it for publication was Xiping Hu.

2) Coverage Bias: A bias where non-uniform space and/or time is given to different news stories.

3) Statement Bias: A bias where specific slants are given to the coverage of a news story.

Gate-keeping bias corresponds to no actual coverage being granted to a specific news story, and hence is a special case of coverage bias. Although not proven or legally agreed upon, media bias is apparent and very common these days. This bias has been there for a long time, e.g., the formal accusation on Fox news about the misinterpretation of facts in hopes of favouring the Conservatives [4]. Media bias not only affects the reputation of the media sources alone, there are other various stakeholders as well such as the general public, the journalists, the regulation agencies or media cells of watchdog organizations that are affected by this bias. Media sources guilty of violating the code of conduct, the anti-discriminatory laws or government censorship are penalized with distribution restrictions or complete closure of the entire news source. It is also considered unethical to be biased or discriminatory towards (or against) minority or socially
sensitive groups [5]. If news sources are exposed for practicing biasness towards or against a certain group, its user ratings and advertisement affiliations can drop drastically, ultimately leading to severe financial losses. Some examples of regulation agencies taking action against news sources include the banning of certain media sources such as “CNN”, “New York Times” and “Guardian” from off-camera press “gaggle” with the White-house press secretary [6], “Channel One” in Lithuania for their propaganda about Ukraine crisis [7] and the suspension of “Geo News” in Pakistan for 15 days with a fine of 10 million PKR (Pakistani Rupees) in June 2014 for its biased coverage of government and military agencies [8]. A naive analysis of media data can produce misleading and biased results. Suppose, for instance, that in a given time period, a source \( S_1 \) generates 200 and 500 tweets respectively, about entities \( E_1 \) and \( E_2 \) while another source \( S_2 \) generates 500 and 600 tweets respectively, for the same settings. A direct comparison indicates that \( S_2 \) has a positive coverage bias for both \( E_1 \) and \( E_2 \). However, this simple comparison disregards the fact that \( S_2 \) has a much higher overall broadcast frequency than \( S_1 \) and, with appropriate rationalization, \( S_2 \) may in fact be biased against \( E_2 \). In this work, our Key Contributions are:

- Presentation of a general model of media bias quantification for statistically consistent measures, intuitive visualizations and ready comparisons, applicable to any set of sources or entities/topics.
- Experimentation on two real-world datasets regarding the general elections in Pakistan, 2018 and the Beirut and Paris bombings in 2015, and showcasing the generality of our approach toward measuring media bias.

The rest of the paper is organized as follows: Section 2 discusses related work. Our model and methodology are presented in Section 3. Section 4 describes the dataset collection procedure along with annotation strategy and Section 5 shows our experimentation and results. Section 6 summarizes our contributions and presents the concluding remarks.

II. RELATED WORK

Prior research related to media bias has many dimensions. Much of the work comes from the social sciences with a smaller body coming from computer science and mathematics. Media bias has remained a well-researched topic during the previous decade. Existing research on media bias, in the domain of social sciences, has employed manual and cumbersome methods of developing bias detection models by including lengthy procedures of data annotations, content analysis and frame analysis etc. However, with the introduction of social media and big data these models should be modified so that they can cope with large amount of news generated by media outlets everyday. Many fast, scalable and automated methods are available in the computer science domain such as, name entity extraction, document search, word co-occurrences, automated sentiment analysis and image processing etc. There is still a lot of room for interdisciplinary research on media bias among social and computer scientists.

Discussion of general and social aspects of media bias includes factors influencing media bias, people’s perspective of bias in media, and influence of news media in elections [9]. The most commonly discussed biases include reporting that supports (or attacks) some entities over the others including political parties, candidates, ideologies, corporations, races, etc [3].

Lin et al. [3] focus on quantifying bias on social media using blogs and news feeds. For their research, they quantify the coverage of US congressmen during 2010 midterm elections to find the political party slants as bias in the news channels. Similarly, research on identification of liberal and conservative bias of news source [9] reveals only the political slant. DiGrazia et al. [10] try to analyse the number of tweets as a metric to represent the number of votes a presidential candidate is likely to receive. The research does not provide a generic model of bias quantification, and justifications are based on presidential elections only. Research on impact of media bias on presidential elections shows that biased reporting by news sources can influence the voting behaviour of the people [4], [11].

Gentzkow and Shapiro [12] state that media bias is often based on the fact that media sources intentionally slant the news reports to be inclined towards the beliefs of their audience, hence giving the impression that news source’s quality and henceforth, authenticity are high. In [13], the authors claim that media outlets express their ideological position in news stories by dis-proportionately criticizing an entity of the opposing ideology. On the contrary, another research [14] predicts that rational citizens are skeptical about the bias in news sources. Therefore, they rely less on the biased sources in their individual decision making. El Ali et al. [15] state that different international news media portray similar events across the globe with varying intensities depending upon the geographical location and social norms of the region such as race, culture and religion. The authors define the intensity as the sympathy of a media source towards a region having experienced an event. Comparison between Arabic and Western media is depicted through a correlation coefficient. Similarly, Gentzkow et al. [2] try to identify the diversity in the opinions of different news media to represent a particular event. The authors claim that this diversity is not because of intrinsic characteristic of the media rather because of the bias of the media towards a specific standpoint. Research on media bias mitigation techniques includes the internet news service called “NewsCube” [16]. This service includes multiple viewpoints regarding an event of interest. It allows the readers to compare and analyse multiple viewpoints regarding an event altogether, thus aiding in the formulation of an unbiased opinion. This reduces the impact of a single biased news producer; hence mitigating the news bias in real time. This solution is more effective than other bias reduction techniques, such as collaborative evaluations for bias reduction [16].
Even though the research does not provide a statistical measure of bias, it provides an automated media bias mitigation solution.

Text Network Analysis/Dynamic Networks deal with network structures based on words as nodes and their co-occurrences as relationships between them. The most influential words form closed communities which can indicate discourse biases. However, our current study differs from this approach and presents a systematic approach to model, quantify and visualize, relative coverage and statement bias. This thorough quantification of media bias can be extremely useful for media managers and regulatory authorities to ensure the fairness of electronic media rather than analyzing the textual semantics of a text network. Recent studies [17], [18] on measuring media bias are also more focused on the use of Text Mining and Natural Language Processing (NLP) techniques. Hamborg et al. focusing on the use of Text Mining and Natural Language Processing (NLP) techniques. Hamborg et al. [17] present an automated approach to identify slanted news coverage by incorporating NLP and deductive content analysis techniques. The research uses the concept of Word Choice Labelling (WCL) for the identification of bias in a single news reported differently by different media sources i.e. using different terminologies for the same event lead to bias. Another work on detection of media bias using text analysis is presented by Paranyushkin [18]. This research focuses on detection and analyzing discourse bias using text network.

Furthermore, studies related to influence propagation in social networks suggest that the nature of a message i.e. its polarity, topic and content can affect the speed, time and volume of information diffusion in these networks. For instance, Henry et al. [19] discuss the volume of negative vs. positive information and its diffusion in social networks like Twitter and Facebook etc. Halberstam and Knight [20] study information diffusion in social networks within groups having homophily. The authors have shown that political information with positive sentiments propagates within like-minded people more quickly than in people with opposing ideologies. Wang et al. [21] combine polarity based information diffusion measures and textual information for detecting sentiment polarities in Twitter. Li et al. [22] studies the information diffusion in signed social networks i.e. the network with nodes having a positive and negative relationship with each other. Particularly, in political relations, where positive relations represent political allies (trust) and negative relations represents political enemies (distrust). However, none of the current influence diffusion models directly deals with quantification and propagation of media bias, especially, both coverage and statement bias. Research related to fairness-aware data mining methods [5], [23], [24] focuses on detection of socially discriminatory practices from a given data set. Proposed fairness-aware data mining methods [25]–[28] prevent undesirable practices from the future model learning processes, however none of these research works deal with media bias and textual data.

### III. METHODOLOGY

Our proposed methodology circulates around a principled approach for understanding and quantifying bias in news media coverage of an event on social platforms such as Twitter. Our proposed model takes as input, tweets on comparable entities or subjects posted by media sources of interest and outputs statistically consistent estimates for bias of each source-entity pair. Our proposed model is general in its application and can be deployed for data in varying formats such as web articles and blogs. For this research, our model has been developed and evaluated for short messages known as tweets.

#### A. MODEL SETTING AND ASSUMPTIONS

Let there be $N$ media sources $S_i (i = 1, \ldots , N)$ that tweet on Twitter regarding $M$ entities or subjects $E_j (j = 1, \ldots , M)$ in a given time period. It is assumed that the entities under consideration are of the same type (e.g., political parties, presidential candidates, technology companies, etc.) and their comparative analysis makes intuitive sense. Similarly, when dealing with subjects or events, it is assumed that the events under study are comparable (e.g. cricket Champion’s Trophy tournament). Our model can be generalized towards situations with:

- A many-to-one relationship between entities/subjects and source.
- A one-to-many relationship between the entity/subject and the sources.

Let $n_{ij} \geq 0$ be the number of tweets published by source $S_i$ on entity/subject $E_j$. Of these tweets, let $n_{ij}^+$ represent the number of tweets with a positive polarity and $n_{ij}^-$ represent the number of tweets with a negative polarity regarding $E_j$ by $S_i$. The polarity rate of $S_i$ for $E_j$ is defined as $p_{ij}$ where:

$$p_{ij} = n_{ij}^+ / (n_{ij}^+ + n_{ij}^-)$$

This information, captured in the contingency table as shown in Table 1, represents our mini-world. The rows and columns of this table represent our sources and entities/subjects respectively. The rightmost cells of each row represent the row sums whereas the bottom cells of each column in the contingency table show the column sums. The accumulative sum of all cells of the contingency table is given in the bottom right cell and is denoted by $n_+$ and $p_+$.

#### TABLE 1. Contingency table showing counts and polarity rates for sources ($S$) and entities ($E$) in our mini-world.

| Sources | $E_1$ | $E_2$ | $E_M$ | Sums |
|---------|-------|-------|-------|------|
| $S_1$   | $n_{11}, p_{11}$ | $n_{12}, p_{12}$ | $n_{1M}, p_{1M}$ | $n_{1+}, p_{1+}$ |
| $S_2$   | $n_{21}, p_{21}$ | $n_{22}, p_{22}$ | $n_{2M}, p_{2M}$ | $n_{2+}, p_{2+}$ |
| $S_M$   | $n_{M1}, p_{M1}$ | $n_{M2}, p_{M2}$ | $n_{M+}, p_{M+}$ | $n_{+}, p_{+}$ |

It is assumed that a consistent approach has been followed to identify tweets regarding a particular entity or subject. One of the possible approaches can be keyword-based
filtering where the keywords are as per the suggestions of domain experts. Another assumption for this research is that a suitable method has been employed for sentiment polarity detection. Given these assumptions, there is a probability that a tweet generated by a source represents multiple entities simultaneously and multiple sources have broadcasted this tweet.

**B. ESTIMATING COVERAGE AND STATEMENT SCORES**

The two key aspects of media bias that have been explored and identified in [4], are coverage and statement. The amount of time and/or space given to a particular entity by a media source is termed as “coverage” whereas the polarity of that coverage is termed as “statement”. For the quantification of these aspects in a mini-world, the following points are to be taken into consideration:

- Various sources may have varying coverage and statement characteristics, for instance, one source may have a much higher tweet frequency than others, or one source may be more biased towards an entity and thus may represent that entity more positively.
- Various entities/subjects may have varying coverage and statement characteristics, for instance, some entities/subjects or topics may inherently be more positive than others.

Taking into consideration the above points, we handle the coverages and statements of the sources and entities with a principled normalization approach. Let \( n_{ij} \) be the coverage of entity \( E_j \) by source \( S_i \) i.e. the number of tweets by \( S_i \) regarding \( E_j \). The coverage precision of \( S_i \) for \( E_j \), calculated by \( n_{ij}/n_i \) (the ratio of the number of tweets published by \( S_i \) about entity \( E_j \) and the total number of tweets by source \( S_i \)). This depicts the relative coverage of \( E_j \) among all entities covered by \( S_i \). Similarly, the coverage recall of \( E_j \) by \( S_i \), calculated as \( n_{ij}/n_j \) (the ratio of the number of tweets about \( E_j \) that have been published by source \( S_i \) and the total number of tweets about \( E_j \)). This depicts the relative coverage of \( E_j \) by \( S_i \) among all sources. Subsequently, a coverage score can be defined as the \( F_1 \)-score of the coverage precision and recall. A similar logic can be applied for statement score by operating on the polarity rates, \( p_{ij} \) i.e. statement precision of \( S_i \) for \( E_j \) is calculated by \( p_{ij}/p_i \), whereas statement recall of \( E_j \) by \( S_i \) is calculated as \( p_{ij}/p_j \) and the statement score is the \( F_1 \)-score of the statement precision and recall.

**Definition 1 (Coverage and Statement Score):** Given the mini-world shown in Table 1 for \( N \) media sources and \( M \) entities, the Coverage Score of source \( S_i \) for entity \( E_j \) is defined as

\[
c_{ij} = \frac{2n_{ij}}{n_i + n_j} \quad (1)
\]

Similarly, the Statement Score of source \( S_i \) for entity \( E_j \) is defined as

\[
s_{ij} = \frac{2p_{ij}}{p_i + p_j} \quad (2)
\]

Here, \( n_{ij} \) and \( p_{ij} \) are the observed count and polarity rate of \( S_i \) and \( E_j \).

The coverage and statement score values are between a range of zero and one. A higher coverage score signifies greater coverage whereas a higher statement score represents a greater positive polarity in the mini-world. Interestingly, the degree of association between \( E_j \) and \( S_i \) with respect to coverage and statement, according to the Bray-Curtis or Sørensen-Dice index [29], [30] is also given by Eqs. 1 and 2, respectively. This implies that the above computed scores also represent the inclination of a source towards an entity with respect to the entity’s overall coverage by the source (coverage score) and the positive treatment of the entity (statement score).

Referring to the example introduced in Section I “Introduction”, the coverage scores of \( S_1 \) and \( S_2 \) for \( E_2 \) are \( c_{12} = 0.56 \) and \( c_{22} = 0.55 \), respectively. This shows that \( S_1 \) has a slightly higher coverage score for \( E_2 \) regardless of having published a lesser number of tweets about \( E_2 \) than \( S_2 \).

**C. NEUTRAL SCORES**

Our discussion up till now leads to the question, “What would be the coverage and statement scores if the mini-world is neutral?” To address this, we need an unbiased model of sources and entities that is in accordance with our assumptions for the mini-world and our methods for computing the coverage and statement scores. A model which based on complete independence between sources and entities, as adopted in chi-square test of independence, is not appropriate for our mini-world where identical tweets can be counted for more than one entity. On the other hand, a model based on the assumption that counts and polarity rates for all source-entity pairs are identical disregards the varying inherent behaviours of the sources and entities.

Our model assumes that the neutral values are the averages of the corresponding row and column averages. This approach takes into consideration that sources and entities may have varying count and polarity rate characteristics. According to this model, the unbiased count and polarity rate of source \( S_i \) and entity \( E_j \) is given by:

\[
\tilde{n}_{ij} = \frac{1}{2} \left( \frac{n_i + n_j}{M} \right) \quad (3)
\]

\[
\tilde{p}_{ij} = \frac{1}{2} \left( \frac{p_i + p_j}{N} \right) \quad (4)
\]

With these values for counts and polarity rates, the unbiased coverage and statement scores are computed in the same manner as discussed in the previous subsection, leading to the following definitions.

**Definition 2 (Neutral Coverage and Statement Score):** Given the mini-world shown in Table 1 for \( N \) media sources and \( M \) entities, the unbiased/neutral Coverage Score of source \( S_i \) for entity \( E_j \) is defined as

\[
\tilde{c}_{ij} = \frac{2\tilde{n}_{ij}}{n_i + \tilde{n}_j} \quad (5)
\]
Similarly, the unbiased Statement Score of source $S_i$ for entity $E_j$ is defined as

$$\tilde{s}_{ij} = \frac{2\bar{p}_{ij}}{\bar{p}_i + \bar{p}_j} \quad (6)$$

Here, $\bar{p}_{ij}$ and $\bar{p}_j$ are the unbiased count and polarity rate of $S_i$ and $E_j$ defined in Eqs. 3 and 4.

Referring to our running example, the unbiased coverage scores of interest are $\bar{c}_{12} = 0.50$ and $\bar{c}_{22} = 0.55$.

**D. ABSOLUTE AND RELATIVE BIAS**

We define two types of measures for assessing media bias: absolute and relative. Absolute bias measures the deviation of observed scores from their unbiased counterparts for a given source-entity pair, whereas relative bias quantifies the difference in score between two sources for a given entity or between two entities by a given source.

**Definition 3 (Absolute Coverage and Statement Bias):**

Given the mini-world shown in Table 1 for $N$ media sources and $M$ entities, the Absolute Coverage Bias and the Absolute Statement Bias of source $S_i$ for entity $E_j$, denoted by $ACB_{ij}$ and $ASB_{ij}$, is defined as

$$ACB_{ij} = c_{ij} - \bar{c}_{ij} \quad (7)$$

$$ASB_{ij} = s_{ij} - \bar{s}_{ij} \quad (8)$$

Here, $c_{ij}$, $s_{ij}$, $\bar{c}_{ij}$, and $\bar{s}_{ij}$ are defined in Eqs. 1, 2, 5, and 6, respectively.

**Definition 4 (Relative Coverage and Statement Bias):**

Given the mini-world shown in Table 1 for $N$ media sources and $M$ entities, the Relative Coverage Bias and the Relative Statement Bias between sources $S_i$ and $S_j$ for entity $E_k$, denoted as $RCB(S_i, S_j | E_k)$ and $RSB(S_i, S_j | E_k)$, is defined as

$$RCB(S_i, S_j | E_k) = c_{ik} - c_{jk} \quad (9)$$

$$RSB(S_i, S_j | E_k) = s_{ik} - s_{jk} \quad (10)$$

Here, $c_{ij}$ and $s_{ij}$ are defined in Eqs. 1 and 2, respectively.

The values of absolute and relative coverage/statement bias vary from $-1$ to $+1$. A negative value signifies negative bias and a positive value signifies positive bias, while the magnitude of the values signifies the strength of the bias.

Referring to our running example, the absolute coverage bias of interest are $ACB_{12} = 0.56 - 0.50 = 0.06$ and $ACB_{22} = 0.55 - 0.55 = 0.00$. This implies that $S_1$ is expressing a significant positive bias with respect to coverage for $E_2$ while $S_2$’s coverage of $E_2$ is unbiased.

The absolute and relative coverage and statement bias, combined with the polarity rates, provide a complete quantification of media bias with respect to a mini-world of interest. These quantities can be visualized intuitively in Cartesian plots as illustrated in the next section.

**IV. TWITTER DATA COLLECTION**

Two different datasets were collected and used in our research. Details of these datasets are described below.

**A. PAKISTAN GENERAL ELECTION 2018**

We collected 15,068 tweets on political leaders of three famous political parties: Imran Khan (PTI), Nawaz Sharif (PMLN) and Bilawal Bhutto (PPP), competing in Pakistan General Election held on 25th July, 2018. Tweets were collected from twitter accounts of six famous news channels (based on their followers count) [31] in Pakistan. These six news sources include: Express Tribune, ARY News, Dunya News, Samaa TV, Dawn News and Geo News respectively. Twitter Search API [32] was used during the data collection process. The reason for opting Twitter as data source lies in the fact that a) Twitter has lately become a popular source of information propagation by news media and b) the short length of tweets made it suitable for labelling and processing tasks as opposed to lengthy articles. Duplicates were removed from the dataset and a total of 2,804 unique tweets were retrieved. Retweets were excluded from the dataset. Data Collection spanned 55 days, from 1-06-2018 till 25-07-2018. Tweet distribution against each of the three entities is shown in Fig. 1.

**B. PARIS AND BEIRUT ATTACK 2015**

Our second dataset was on Paris and Beirut Attacks of 2015 [15]. The dataset consists of Western and Arab media tweets about Paris and Beirut attacks (November 2015) and is available at the author’s github account. It consists of tweet IDs, human annotated sentiment scores and other metadata in 4 different languages (Arabic, English, Spanish and French). We only include English language tweets in our dataset to avoid the language complications. A Fleiss Kappa [33] of $\kappa = 0.27$ is reported for English tweets in the dataset. $N = 1,732$ english tweets IDs were present in the original dataset, out of which only 722 tweets were successfully retrieved. The reduction in the number of retrieved tweets is obvious since some tweets and Twitter accounts were removed either by Twitter or by the users themselves.

**C. DATA EXTRACTION AND PRE-PROCESSING**

Our first step was to identify influential news sources and political leaders’ on Twitter. The list of news media sources chosen and number of tweets collected are shown in Table 2. To ensure that news sources chosen are popular among audience, we select six news channels based on their follower counts from a website SocialBakers [31]. It can be noticed that there exist no benchmark or ranking for popularity of news media on Twitter except for the follower count. Similarly, popular political parties and their leaders are manually chosen, we also cross checked our selection using public lists (e.g. Wikipedia” and “The News” article). Twitter Search API is used to retrieve all tweets

---

1https://github.com/abdoelali/CrisisNewsSympathy; last retrieved: 10-10-2018

2https://en.wikipedia.org/wiki/2018_Pakistani_general_election; last retrieved: 11-10-2019

3https://goo.gl/MedPri; last retrieved: 11-10-2019
from these accounts from 1-06-2018 to 25-07-2018 (Election Day). A total of $N = 15,068$ tweets were collected from all six news media sources. Hashtags are extracted from all of these tweets and hashtags against popular political leaders are selected based on their frequency of occurrence. The dataset of ($N = 15,068$) tweets is then filtered based on these hashtags resulting in $N = 2,804$ tweets. Individual distribution of these tweets against each news channel can be seen in Table 2. Further pre-processing includes - converting tweet text to lower case letters, removing mentions (@username), URLs and punctuation marks removal.

**TABLE 2. Fleiss Kappa for different News media.**

| Sr. no | Name of Source | # of Tweets | Fleiss Kappa ($\kappa$) |
|--------|----------------|-------------|-------------------------|
| 1      | Express Tribune | 306         | 0.328                   |
| 2      | ARY News        | 457         | 0.717                   |
| 3      | Dunya News      | 923         | 0.328                   |
| 4      | Samaa TV        | 225         | 0.400                   |
| 5      | Dawn News       | 515         | 0.438                   |
| 6      | Geo News        | 378         | 0.243                   |

**D. SENTIMENT ANALYSIS USING LEXICON AND HUMAN ANNOTATORS**

To calculate statement bias from our dataset, we needed polarity count of the given tweets. Since, there is no known ground truth for the polarities of this specific dataset, we have employed two strategies for polarity calculations of tweets.

1) LEXICON BASED METHOD

Many lexicon based methods are present in the literature which can predict sentiment on short text [34], [35]. Vader [36] is one such open source lexicon, that we have used for the polarity detection of tweets in our datasets. Vader enjoys unchallenged popularity when it comes to the sentimental analysis on social media text along with other lexical based methods like SentiStrength and Afinn. However, our reason for using Vader is its open source code and good performance on benchmark Twitter datasets [37].

2) HUMAN EVALUATION

To gain confidence on our lexicon based polarity measures, we had our Pakistan general election dataset annotated. The dataset was divided into two sets of three channels each. Each set contained almost equal number of tweets (set1 = 1,350 tweets, set2 = 1,454 tweets) and annotated by three different annotators. We made sure that the annotators are well versed in the subject and language of tweets. Inter-rater reliability measure of Fleiss Kappa ($\kappa$) was used to assess the agreement between annotators. The kappa statistic $\kappa$ can be defined as:

$$\kappa = \frac{P_o - P_e}{1 - P_e}. \quad (11)$$

Here, $P_o$ is the observed overall agreement, and $P_e$ is the expected mean proportion of agreement due to chance. If $\kappa = 1$ then there is a complete agreement between annotators. If $\kappa \leq 1$ then there is no agreement at all. Hence, a total of 6 annotators were deployed and an average Fleiss kappa of $\kappa = 0.409$ is reported, which is considered a good agreement score. Highest Fleiss kappa is reported for ARY News with a value of ($\kappa = 0.717$). Detailed kappa statistics against each news source can be seen in Table 2. Next, the final tweet labels were selected by a majority voting algorithm. For each entity, average polarities were measured from both
lexicon and calculated tweet labels, and an F-score of 0.56 is obtained.

V. EXPERIMENTS AND RESULTS

In this section, we present the experimental results demonstrating the generality of our model, the ease of its applicability, and the effectiveness of the proposed visualizations. We focus on the analyses of two datasets: Pakistan general elections 2018 and bombings in Beirut and Paris 2015.

A. PAKISTAN GENERAL ELECTIONS 2018

We analyze media bias in the coverage of Imran Khan ($E_1$), Nawaz Sharif ($E_2$) and Bilawal Bhutto ($E_3$) for the Pakistan’s general elections of 2018. For this purpose, we collected tweets from the official channels of 6 news media sources as explained in section IV-A. The polarity of the gathered tweets is calculated using the method explained in IV-D.1. The key characteristics of this data are given in Table 3.

| No. | News Sources     | Counts | Polarity Rates |
|-----|------------------|--------|----------------|
| S1  | Express Tribune | 29     | 0.734 0.713    |
| S2  | ARY News         | 100    | 0.740 0.655    |
| S3  | Dunya News       | 88     | 0.742 0.708    |
| S4  | Samaa TV         | 27     | 0.765 0.712    |
| S5  | Dawn News        | 71     | 0.808 0.636    |
| S6  | Geo News         | 60     | 0.676 0.648    |

1) RELATIVE BIAS

In Figures 2, 3 and 4 the relative coverage and statement bias between sources for leaders of each of the famous Pakistani political parties has been plotted. The figures show the relative scores calculated by subtracting the absolute scores of a source from the others and are represented in bars.

From Fig. 2 it is observed that $S_2$ (ARY news) has a much higher coverage of Bilawal Bhutto (see the blue bar) (i.e., Positive relative coverage bias) while $S_6$ (GEO News) has a much lower coverage of Bilawal Bhutto (i.e., Negative relative coverage bias). However, in terms of relative statement bias, $S_6$ (GEO News) is always positively biased (see the green bar) in statements about Bilawal Bhutto than any other source. For Imran Khan, in Fig. 3 it can be seen that $S_3$ (Dunya News) has a higher coverage than any of the other sources while $S_1$ (Express Tribune) has a much lower coverage as
comparing to the rest. This hints at the fact that $S_1$ (Dunya News) gives a more positive exposure to Imran Khan while $S_1$ (Express Tribune) gives him a lesser exposure in comparison to the other sources. The inter-source relative statement bias for Imran Khan is also shown in Fig. 3. Here, it is observed that $S_3$ (Dawn news) is more positive in its statements about Imran Khan (see red bar in relative statement score) than any other news source, while $S_4$ (Samaa) has a high negative statement bias towards Imran Khan. In Fig. 4, negative scores for relative coverage bias towards Nawaz Sharif by source $S_1$ (Express Tribune) can be clearly seen. Another interesting observation shown in Fig. 4 is that $S_3$ (Dunya news) has highest positive relative statement and coverage bias for Nawaz Sharif in comparison with other sources. This indicates that $S_3$ (Dunya news) is more biased towards Nawaz Sharif, hence giving him a more positive coverage as opposed to the rest.

These visualizations while demonstrating the ease with which relative bias of multiple sources can be understood also highlight certain patterns that would have been too difficult to discern otherwise. This gives a clear indication of the relative statement and coverage biases that news media practice for political parties. This is extremely important for the understanding of information dissemination to the common man.

2) ABSOLUTE BIAS

We developed a visualization capturing the absolute coverage and statement bias in a mini-world, together with the polarity rate of each source-entity pair, in a comprehensive and intuitive manner. Fig. 5 shows this visualization for the leaders of the mainstream political parties of Pakistan. In this Fig., each source-entity pair is represented by a shape in a 2-D space where $x$-axis is the absolute statement bias and $y$-axis is the absolute coverage bias. For ease of comprehension, a particular shape depicts a particular news source and the colour of the shape depicts the entity under consideration (Imran Khan, Nawaz Sharif and Bilawal Bhutto).

![FIGURE 5. Absolute bias for political leaders of top political parties.](image1)

This visualization clearly depicts the bias characteristics of the media sources. In particular, the source-entity pairs in the upper right quadrant (e.g., Dunya News - Nawaz Sharif) exhibit positive bias while source-entity pairs in the lower left quadrant (e.g., Dunya News - Bilawal Bhutto) express negative bias. In terms of coverage, Dunya news exhibits a strong positive bias towards Nawaz Sharif while Dawn news expresses the same for Imran Khan. In terms of statement, we conclude that while other sources have mixed statement scores i.e., both negative and positive for Bilawal Bhutto and Imran Khan, they all have strong negative statement bias for Nawaz Sharif.

B. PARIS AND BEIRUT ATTACKS 2015

In this experiment, we quantify and visualize the bias in news sources towards the 2015 bombings in the Middle East and Table 4 shows the mini-world indicating tweet counts and polarity rates for Paris and Beirut bombings. From Fig. 6 it can be clearly seen that while Arab media had a higher coverage and statement bias towards Beirut than Western media (which is acceptable since it is a regional incident for Arab community), their coverage of the Paris Attacks is comparable to that of the Western media channels as well i.e., the two have a rather similar coverage and statement bias towards the incident. On the other hand, it can be clearly seen that Western media had polar opposite reactions towards both incidents as one of their points lies in the upper right quadrant (positive coverage and
statement bias) while the other lies in the lower left quadrant (negative coverage and statement bias). These findings are in line with [15] where they concluded that western media had lower coverage and sympathy towards the Beirut Attacks while Arab media had comparably higher coverage and sympathy for Paris attacks. However, the authors didn’t quantify bias in the mentioned paper, rather, they calculated the temporal correlation in their study. In our research, we have clearly presented these findings through our bias calculations and insightful visualization.

VI. CONCLUSION

We present a statistically consistent and generic model for quantifying media bias towards different personalities and topics. Our model presents a ready to use approach where with slight modifications, it can be used to calculate biases in a variety of datasets, be it micro-blogs or online forums. We have shown our model at work by calculating the bias measure for six top news sources in Pakistan towards the chairpersons of the top three political parties in Pakistan. Our detailed empirical study shows a negative bias towards Nawaz Sharif by the majority of news media sources. We also validate our proposed approach on a second dataset of Paris and Beirut attacks of 2015 and our results are in line with [15]. Moreover, our visualizations and quantification give a better perception of the relative bias shown by international media.

REFERENCES

[1] P. J. Shoemaker and S. D. Reese, Mediating Message 21st Century: A Media Sociology Perspective. Evanston, IL, USA: Routledge, 2013.
[2] M. Gentzkow, J. M. Shapiro, and D. F. Stone, “Media bias in the marketplace: Theory,” in Handbook of Media Economics, vol. 1. Amsterdam, The Netherlands: Elsevier, 2015, pp. 623–645.
[3] Y. R. Lin, J. P. Bagrow, and D. Lazer, “More voices than ever? Quantifying media bias in networks.” 2011, arXiv:1111.1227. [Online]. Available: http://arxiv.org/abs/1111.1227
[4] D. Saez-Trumper, C. Castillo, and M. Lalmas, “Social media news communities: Gatekeeping, coverage, and statement bias,” in Proc. 22nd ACM Int. Conf. Inf. Knowl. Manage. (CIKM), 2013, pp. 1679–1684.
[5] D. Pedreschi, S. Ruggieri, and F. Turini, “Discrimination-aware data mining,” in Proc. 14th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), 2008, pp. 560–568.
[6] L. Pasha-Robinson. (2017). White House Media Ban is ‘Unconstitutional’ by a trusted source,’’ Amer. J. Political Sci., vol. 44, no. 2, pp. 301–315, Apr. 2000.
[7] T. M. Times. (2015). Lithuania May Block Two More Russian Tv Channels, [Online]. Available: http://www.themoscowtimes.com/business/article/lithuania-may-block-two-more-russian-tv-channels-for-biased-coverage/514084.html
[8] T. E. Tribune. (2014). Pentra Suspends Geo Tv’s License for 15 Days, Imposes Rupees 10 Million Fine, [Online]. Available: http://tribune.com.pk/story/117288/pentra-suspends-geo-tvs-license-for-15-day-s-imposes-rs10-million-fine
[9] J. M. Miller and J. A. Krosnick, “News media impact on the ingredient of presidential evaluations: Politically knowledgeable citizens are guided by a trusted source,” Amer. J. Political Sci., vol. 44, no. 2, pp. 301–315, Apr. 2000.
[10] J. DiGrazia, K. McKeelv, J. Bollen, and F. Rojas, “More tweets, more votes: Social media as a quantitative indicator of political behavior,” PloS ONE, vol. 8, no. 11, 2013, Art. no. e79449.
[11] S. DellaVigna and E. Kaplan, “The fox news effect: Media bias and voting,” Quart. J. Econ., vol. 122, no. 3, pp. 1187–1234, Aug. 2007.
ANAM ZAHID (Member, IEEE) received the M.S. degree in information technology from the National University of Sciences and Technology (NUST), Islamabad, in 2015. Her thesis work was based on proposing and implementing security in NoSQL databases. As part of her thesis, she applied encryption and fine-grained access control in the sharded architecture of NoSQL database MongoDB. She has served as a Lecturer for two years with the Computer Science Department, National Textile University (NTU), Faisalabad. She has also published one conference and one journal article in the prestigious venues of IEEE and Elsevier prior to her Ph.D. She is currently a Ph.D. Fellow with the Data Science Lab (DSL), Information Technology University (ITU), under the supervision of Dr. F. Kamiran. Her current area of research is ethical issues in social networks, particularly topics related to detection and mitigation of fake news, unfair popularity, misinformation, and media bias.

MAHAM NASIR KHAN received the bachelor’s degree in electrical engineering from the University of Engineering and Technology, Lahore, in 2017. She is currently pursuing the master’s degree in data sciences with Information Technology University, Lahore. She has also been associated with the Data Science Lab, ITU, as a Graduate Fellow, since 2019. Her research interests revolve around data mining, data visualization, social media analysis, machine learning, and deep learning.

AHMER LATIF KHAN received the bachelor’s degree from the Pakistan Institute of Engineering and Applied Sciences, Pakistan, in 2014, and the M.S. degree in electrical engineering from the Lahore University of Management Sciences, Pakistan. He is currently a full-time Research Associate with the Data Science Lab (DSL), Information Technology University, Pakistan. He has interdisciplinary research interests include spanning network security, data science, and artificial intelligence. Particularly, he works on composable research problems that are common across multiple domains. His current research projects investigate the dynamics of social networks, with an emphasis on its discriminatory dimension and ethical use.

FAISAL KAMIRAN received the Ph.D. degree from the Eindhoven University of Technology (TU/e), The Netherlands, in 2011. Afterwards, he joined the King Abdullah University of Science and Technology (KAUST) as a Postdoctoral Fellow. As a part of his Ph.D. research, he developed fairness-aware AI and machine learning models, which are directly applicable to the legal, social, and economic domains particularly after strict privacy regulations like GDPR. The Dutch Department of Justice (WODC) and The Dutch Central Bureau of Statistic (CBS) have applied his developed techniques to avoid racial and gender discrimination in the models they develop to support the decision-makers in The Netherlands. He is currently an Assistant Professor with Information Technology University. He is also the Chairperson of the Computer Science Department, the Dean of Engineering, and the Director of the Data Science Lab (DSL), Information Technology University. He is also the Co-Founder and the President of ADDO AI. He received the World Bank Research Innovation Award on his data science work, in April 2017. His research interests include fairness-aware data analytics and machine learning to safeguard the rights of the deprived communities and individuals, ICTD, social media analytics, text mining, and so on. He has published more than 30 refereed articles in various prestigious journals and conferences. His publication list includes several research articles in the prestigious A*/A data science venues, such as ICDM, ECML/PKDD, EMNLP, PAKDD, CIKM, SAC, and KAIS. According to Google Scholar, his H-index is 10, and his research work has more than 650 citations.

BILAL NASIR received the B.S. degree in computer information technology with specialization in database management from Purdue University, IN, USA, in 2004, the first M.S. degree in strategic IT management from Stockholm University, Sweden, in 2010, and the second M.S. degree in computer science from Information Technology University, Lahore, in 2015. His M.S. research thesis was in the domain of media bias quantification in social networks. Moreover, he had wide experience in working on diverse job roles, including a Software Engineer, a Business Analyst, a Database Administrator, a Data Modeler, and a Data Scientist. His current research interests include social media bias identification, predictive algorithms, text mining, and machine learning.