bias goggles: Exploring the Bias of Web Domains through the Eyes of Users

(demo of the ECIR 2020 paper)

G. Konstantakis\textsuperscript{2}, G. Promponas\textsuperscript{2}, M. Dretakis\textsuperscript{2} and \textsuperscript{2} P. Papadakos\textsuperscript{1,2}

\textsuperscript{1} Institute of Computer Science, FORTH-ICS
\textsuperscript{2} Computer Science Department, University of Crete
GREECE
Outline

• Motivation

• The bias goggles Model

• Graph-based Computation of Bias

• Evaluation Results

• Demo

• Conclusion & Future Work
Motivation
Transparency, disinformation, and bias are in the focus of our information society.

There is a selective and restrictive exposure to online information:
- Echo chambers and filter bubbles
- Isolation of groups of people

This is a result of our human nature (e.g., Psychology, Sociology)

...reinforced by algorithms:
- Personalization in social networks and WSE
- Biased-data in deep-learning and ML or biased algorithms

This isolation might inhibit the growth of informed and responsible humans/citizens/consumers, and can promote and resurrect social, religious, ethnic, and other kinds of discrimination and stereotypes.
Focus is on the **transparency, fairness, accountability** and **bias** of ML algorithms

- for decision making and recommendation tasks
- ... and for social platforms

However, social platforms mainly act as gateways to information published on the web as common everyday web pages (e.g., blogs and news)

- ... and users are not aware of the bias characteristics of the information they consume

**There is no formal definition of what bias is**
We present the bias goggle model, where users are able to explore the bias characteristics of the web domains for specific user-defined biased concepts (BCs) and aspects of bias (ABs).

We present graph-based algorithms for computing the bias score using popular propagation models and the new Biased-PageRank that models a biased surfer.

We discuss some preliminary evaluation results.

We demonstrate the approach.
Model
Biased Concept (BC)
- the **goggle** through which the user explores the web about bias characteristics
- corresponds to a concept that can be biased (e.g., the politics in Greece)
- user defined
- each BC has at least **two ABs**

Aspect of Bias (AB)
- corresponds to **a bias dimension for a BC** (e.g., a specific political party)
- ABs are **user defined** through a non-empty set of **seeds** (i.e., domains) that the user considers to fully support this bias dimension
- A seed can be used in only one **AB** of a specific **BC**

Both AB and BC are identified by **SHA-1** signatures over their data
Incomparable Seed Assumption

The **domains** in the **set of seeds** of an AB are **incomparable** and **equally supportive** of this AB

- e.g., the homepage of a political party and the homepage of its leader are assumed incomparable

Orthogonality of Aspects of Bias

All **ABs** in any user-defined BC are **considered orthogonal**

- e.g., the ABs related to the political parties in the parliament are considered orthogonal and unrelated
The support of domain to an AB $\text{sup}(AB, dom)$ is a measure of \textit{“supportiveness”} of the web domain to a specific AB

- A core metric for our model
- Normalized score in $[0, 1]$

Various approaches \textbf{to compute this score} over a set of web pages

- the \textbf{graph-based} ones that exploit the web graph structure and the relationship of a domain with the domains in the graph (\textbf{our focus})
- \textbf{content-based} approaches that consider the textual information of the respective web pages
- \textbf{hybrid} approaches that exploit both
The support of a domain to a BC $\text{sup}(BC, A)$ is a measure of the “supportiveness” of the web domain to a specific BC

- It depends on the support scores of the distinct ABs of this BC
- Normalized score in [0, 1]

It is the norm of the vector of the support scores of the ABs Normalize it by dividing with the ideal supportive vector

- The ideal supportive vector has $\text{sup}(AB, \text{dom}) = 1$ for all ABs of a BC
The **bias score of a domain to a BC** captures how biased a domain is to a BC

- Normalized score in $[0, 1]$

It depends on

- the **support score of the BC** for this domain (i.e. highly biased domains should be highly supportive of the BC)

- the **distance of the support vector** from the non-biased unit vector that fully supports all ABs

\[
\text{bias}(s_{\text{dom}}^{d_{A}}) = \text{dist}(s_{\text{dom}}^{d_{A}}, 1_{|A|}) \times \text{sup}(BC_A, \text{dom})
\]

\[
\text{dist}(s_{\text{dom}}^{d_{A}}, 1_{|A|}) = 1 - \cos \text{Sim}(s_{\text{dom}}^{d_{A}}, 1_{|A|}) = 1 - \frac{s_{\text{dom}}^{d_{A}} \cdot 1_{|A|}}{\|s_{\text{dom}}^{d_{A}}\| \|1_{|A|}\|}
\]
Graph-based Computation of Support Scores of ABs
Let $W$ be a set of crawled pages

$\text{doms}(W)$ the set of normalized SLD or TLD level domains in $W$ (e.g. ecir2020.org)

$\text{links}(W)$ holds all the crawled links between the domains in $W$

$G(W)$ the corresponding graph with nodes the $\text{dom}(W)$ and edges the $\text{links}(W)$

The Support Flow Graph $\text{SFG}(W)$ of a set of web pages $W$ is the directed weighted graph that is created by inverting the links in $G(W)$
Equally Supportive Links

- Any link $\text{link}(\text{dom}, \text{dom'})$ in $W$, from the domain $\text{dom}$ to the domain $\text{dom'}$ in the set of crawled domains $W$, is considered to be of supportive nature (i.e., $\text{dom}$ has the same support stance as $\text{dom'}$ for any AB).

- All links from $\text{dom}$ to $\text{dom'}$ are equally supportive and independent of the importance of the page they appear in!

The **weight** of an SFG’ edge$(\text{dom}, \text{dom'})$ is the **number of links** from pages in $\text{dom'}$ to pages in $\text{dom}$, divided by the total incoming number of links to $\text{dom}$ from all domains and takes a value in $[0, 1]$. 
Computing the Support Scores

Given an $SFG(W)$ and the seeds of an AB we can now describe how the support flows in the nodes of the SFG(W) graph for each AB

- **Independence Cascade** Propagation
- **Linear Threshold** Propagation
- **Biased-PageRank**
**Basic Algorithm**

1. We run $n$ experiments
2. Each run **starts with the set of activated** nodes (in our case **the seeds of the AB**)
3. In each iteration there is a **history independent and non-symmetric probability** of flowing the support to the neighbors of the activated nodes through their link
4. The **probability** is represented by the **weights** of the **links** in the $SFG(W)$
5. The **nodes** and their **neighbors** are **selected in arbitrary order**
6. Each experiment **stops** when there are **no other activated nodes**
7. After $n$ runs we **compute the average support scores of nodes** which is the final **support score for this AB**
Linear Threshold (LT)

Linear Threshold is similar to Independence Cascade
But for a node to become **active** the sum of the support of its neighbors is **greater** than a **threshold**
  - Resistance of a node
  - This threshold is computed randomly in each experiment
  - Again, we use the probabilities represented by the weights of the SFG links

After **n** runs we **compute the average support scores of nodes which is the final support score for this AB**

**More expensive than IC** since we have to always consider the neighborhood (can be rather large for the Web Graph)
Variation of PageRank that models a biased surfer

The biased surfer always starts from the biased domains (seeds) of the AB and either visits a domain linked by the selected seeds or any of the seeds again

• Based on some probability

We have modeled three behaviors

• Strongly Supportive (SS), where the probability of visiting a seed or the linked domains is the same

• Decreasingly Supportive (DS), where the probability of visiting the linked domains increases as the surfer visits more pages

• Non-Supportive (NS), modeling a surfer that is biased only initially

When the surfer stops visiting links/seeds

• Biased teleportation - higher probability to pages that are close to the seeds

Initially all support scores are 0 except the scores of the seeds with score $1 / |\text{seeds}|$

Run until convergence
Evaluation
We crawled part of the greek web based on the following seeds for crawling:

- 383 domains related to the greek political life
- 89 sport related
- 127 big companies
- top-300 popular greek sites from Alexa

The created SFG contains 90,419 domains and 288,740 links.

We created two BCs:

- One containing all greek political parties in the parliament (9 ABs)
- One containing popular greek football sport teams (6 ABs)

We also manually constructed a golden collection that holds highly supportive domains for each AB of the two BCs from the crawled dataset.

- There are no datasets that associate domains with a degree of bias
- So we comparatively evaluate our algorithms
Average Golden Bias Ratio (AGBR)
The ratio of the average computed bias score of the golden domains divided by the average score of all domains in the dataset
  • The higher the value the easier we can discriminate the golden domains from the rest domains in the dataset for a BC

Average Golden Similarity (AGS)
The average similarity of the computed support vector of the golden domains to their corresponding supportive AB
  • The higher the value the closer the golden domains are found to be towards their supportive AB
Manual inspection shows that some domains in the golden dataset get rather low bias scores
  • This is due to the fact that at least in our dataset they do not contain links to domains in the neighborhood of the seeds

<table>
<thead>
<tr>
<th>Alg.</th>
<th>n</th>
<th>bias</th>
<th></th>
<th></th>
<th>bias</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>t (s)</td>
<td>AGBR</td>
<td>AGS</td>
<td>t (s)</td>
<td>AGBR</td>
<td>AGS</td>
</tr>
<tr>
<td>IC</td>
<td>1000</td>
<td>0.399</td>
<td>217.571</td>
<td>0.2619</td>
<td>0.287</td>
<td>320.408</td>
<td>0.1362</td>
</tr>
<tr>
<td></td>
<td>k/2</td>
<td>0.622</td>
<td>230.798</td>
<td>0.3024</td>
<td>0.497</td>
<td>326.005</td>
<td>0.1365</td>
</tr>
<tr>
<td></td>
<td>k</td>
<td>0.945</td>
<td><strong>234.535</strong></td>
<td>0.361</td>
<td>0.681</td>
<td>328.667</td>
<td>0.1675</td>
</tr>
<tr>
<td>LT</td>
<td>1000</td>
<td>129.0</td>
<td>230.611</td>
<td>0.2354</td>
<td>87.5</td>
<td>322.585</td>
<td>0.1528</td>
</tr>
<tr>
<td></td>
<td>k/2</td>
<td>512.9</td>
<td>230.064</td>
<td>0.3621</td>
<td>322.5</td>
<td><strong>328.698</strong></td>
<td>0.1364</td>
</tr>
<tr>
<td></td>
<td>k</td>
<td>966.1</td>
<td>231.087</td>
<td>0.3626</td>
<td>663.1</td>
<td>327.786</td>
<td>0.1659</td>
</tr>
<tr>
<td>Biased-PR</td>
<td>SS  (40, 31)</td>
<td>34.8</td>
<td>227.999</td>
<td><strong>0.5569</strong></td>
<td>17.9</td>
<td>261.788</td>
<td><strong>0.4745</strong></td>
</tr>
<tr>
<td></td>
<td>DSS (319, 391)</td>
<td>260.5</td>
<td>129.829</td>
<td>0.3730</td>
<td>219.4</td>
<td>163.165</td>
<td>0.4344</td>
</tr>
<tr>
<td></td>
<td>NSS (306, 458)</td>
<td>231.4</td>
<td>32.602</td>
<td>0.3041</td>
<td>207.6</td>
<td>34.905</td>
<td>0.4052</td>
</tr>
<tr>
<td></td>
<td>BC1 - Political Parties (6 AEs)</td>
<td>129.0</td>
<td>230.611</td>
<td>0.2354</td>
<td>87.5</td>
<td>322.585</td>
<td>0.1528</td>
</tr>
<tr>
<td></td>
<td>BC2 - Sports Teams (6 AEs)</td>
<td>322.5</td>
<td><strong>328.698</strong></td>
<td>0.1364</td>
<td>663.1</td>
<td>327.786</td>
<td>0.1659</td>
</tr>
</tbody>
</table>

Results
Introduction the bias goggles model

- Exploration of the bias characteristics of web domains
- ... to user-defined biased concepts and aspects of bias

Graph-based approaches for computing the bias score

- using popular propagation models
- the new Biased-PR PageRank variation that models biased surfers behaviors

Implementation of a first prototype

A comparative evaluation over a subset of the greek web shows promising results

A manually constructed dataset of biased concepts and biased domains
Future Work

Variations with **more relaxed assumptions** about our model
- e.g., *supportive, neutral or opposite* nature of **links using sentiment analysis**
- **importance** of the web pages they appear in

More experiments with **graph-based** approaches

**Content-based** approaches
- topic extraction algorithms (e.g., LDA)

Exploitation of other **social network graphs** (e.g., retweet, friends)

Implementation of **scalable algorithms and indexes**
- Crawl more data – we need **millions** of **popular** domains
- **Optimize speed of support computation** for real-time needs (e.g. GPU)
- Incremental algorithms
Future Work

- **Release the browser plugins** so that they can be used by everyday users
- **Large scale user study** in cooperation with psychologists/sociologists
  - How do users define bias?
  - Which are the most **popular** Biased Concepts?
  - Is there any **correlation** between Biased Concepts of different domains?
  - How is the **behavior** of users affected through time when they are aware of the bias score of the information they consume?
- **Extend our golden dataset** to include more data and data of international interest
Thank you for your attention!

bias goggles
http://pangaia.ics.forth.gr/bias-goggles/

Contact: Panagiotis Papadakos (papadako@ics.forth.gr)