




Finding Bias in Political News and Blog Websites

[Link](#) of paper

Group-11:

- Anvesh Kiran
- Varada Raghunath Sai
Pranav

Abstract of the paper

- 
- News and blog websites-political bias (Republican and Democratic)
 - Done by using a Bipartite graph of memes and websites
 - Nodes: Memes and Websites
 - Edges: If a meme(text/ sentence) used in a website, both these are connected by an edge.
 - In this network, bias of a node is calculated from its neighbours bias.

Data Collection :

- We got the meme-website data (memes present in a website) from memetracker.com as suggested by the research paper.
- But, the website-label data (labelling of some websites) is found from another reference paper.
- This data is about U.S elections.

PreProcessing :


- We cleaned the URL's of websites and then labelled them.
- Only then were the websites we got from two different sources matched.

Conversion into Bipartite Graph :

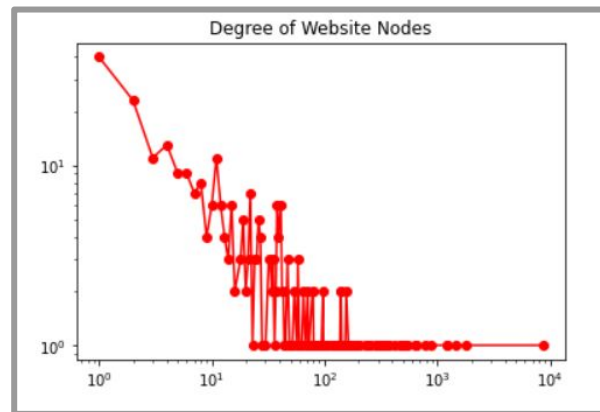
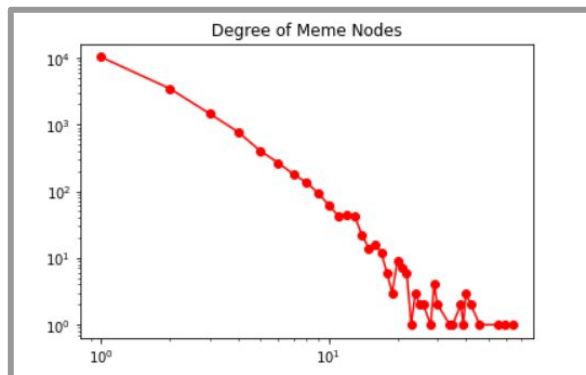
- From the meme-website data, we made the bipartite graph putting an edge between meme and website if the meme is used in it.
- And then from the website-label data, we matched some of the websites and labelled them in the graph as democratic or republican.

Algorithm:

```
1:  $t = 1$ 
2: while Num nodes in  $V_w$  that change bias  $> \epsilon$  do
3:   if  $t$  is odd then
4:      $V_t = V_m$ 
5:   end if
6:   if  $t$  is even then
7:      $V_t = V_w - V_w^l$ 
8:   end if
9:   for all  $v_i \in V_t$  do
10:     $w_{dem}(v_i) = \sum_{v_j \in N_{gh}(v_i)} w_{dem}(v_j)$ 
11:     $w_{rep}(v_i) = \sum_{v_j \in N_{gh}(v_i)} w_{rep}(v_j)$ 
12:     $w_{dem}(v_i) = \frac{w_{dem}(v_i)}{w_{dem}(v_i) + w_{rep}(v_i)}$ 
13:     $w_{rep}(v_i) = \frac{w_{rep}(v_i)}{w_{dem}(v_i) + w_{rep}(v_i)}$ 
14:   end for
15:    $t = t + 1$ 
16: end while
```

- 
- We performed a k-fold cross validation of predicting bias of websites nodes for calculating the accuracy.
 - We also implemented SVM using bag of words and SVM using collection of memes.
 - Our Algorithm accuracy=99.84
 - Acc of SVM using,
 - Bag of words= 98.02
 - Collection of memes=98.4

Results :



Top democratic websites predicted by our algorithm

- demonlover05.deviantart.com
- carafaye.deviantart.com
- thriss7.deviantart.com
- karluwa.deviantart.com
- raga-redkitty.deviantart.com
- kasai-ookami59.deviantart.com
- ozandvm.deviantart.com
- dwk61.deviantart.com
- ubimarble.deviantart.com
- psychotic-butterfly.deviantart.com

Top republic websites predicted by our algorithm

- ebonybird.com
- yoursuccessfulinternetsystems.wordpress.com
- sajpsiterushsite.com
- derfwadmanor.blogspot.com
- vijaysappani.com
- reform-dem.blogspot.com
- dyadeblog.wordpress.com
- offthebaseline.com
- madisonrecord.com
- yellowromancandles.wordpress.com

Top democratic memes predicted by our algorithm

we were called at 10 11pm gmt 9 11am aedt on the 15th to a disturbance at the o2 arena site
given the severity of what is now happening in the global economies we do not believe we will
in this hour of crisis divisive politics will lead us nowhere
nothing can ever justify violence by a person in an intimate relationship
of the 37 million americans living in poverty 27 million are men and children
i'll sign the papers if you will
we are in peace and what they did they killed muslims
at least 10 are injured 10 15 20 are injured i don't know
this was a high-risk situation as soon as they made contact there was gunfire the officers were
and in each case i watched us within a short time focus on new ideas and new solutions and wit

Top republican memes predicted by our algorithm

some may ask how at this moment of economic challenge we can afford to invest in reforming our health
mr hamdan is not an al-qaeda warrior he is not al-qaeda's last line of defence
the united states is reaching out to its partners as well together the steps that we and other men
there's also some other damage on the property caused by wind and water we're in the process of assessing
given the extraordinary budget challenges facing our state and my commitment to continuing the process
we are not aware of any investigation that is under way nor have we been contacted by any agency
it's a show of the kremlin irritation about the u s deployment to georgia it's a signal to the united states
i didn't want to do that and i don't think anybody i know wanted to do that
and these guys are not only rich but they've screwed up
i think leon panetta is totally qualified for this job he's been a consumer of intelligence for a

Conclusion:

- The Accuracy for finding the bias is high for network based approach rather than SVM implementation.
- **Power law** was observed in the above graphs
- **Hubs** are observed
- **Rich become richer and Poor become poor** phenomenon is observed since the algorithm iterates all over the graph by edges.
- Blogs and Websites constitutes more than expected for the political campaign.
- So, Labelling of websites on political bias is highly required.