

Evaluating Twitter’s Algorithmic Amplification of Low-Credibility Content: An Observational Study - Supplementary Materials

1 Baseline Stratification - Quantile-based Discretization

To facilitate the analysis in this study, it was necessary to reduce the overall complexity of the data by clustering the stratification variables into well-defined and comparable bins capturing different clusters for engagement and followers counts profiles. For this purpose, this study relies on quantile-based discretization, an approach that divides the data into similar-sized buckets based on existing quantiles. For the discretization of engagement and followers data, it was observed that by leaving the number of bins free to vary, this converged at 4 bins, which can be considered a division into four clusters, one with the lowest values, (bottom 25%), one with medium values (25 to 50%), one with high values (50 to 75%) and finally a cluster with very high values (75 to 100%). In order to minimise the impact of outliers while binning the data, follower counts and total engagement were initially normalised using a logarithmic scale, which ensures that the bins created are less skewed by outliers. Nonetheless, as it can be expected, the distribution following traditional social media power laws, where the bin containing very high values has a larger range of values. This was not considered concerning, particularly as the use of bootstrapping with a high number of iteration should minimise any such imbalances in the data. Detailed information on the characteristics of each cluster can be found in Tables 1 to 4 below.

Table 1: Clustering - Climate Low Credibility

followers_cluster	engagement_cluster	engagement_min	engagement_max	engagement_mean	engagement_median	impressions_min	impressions_max	impressions_mean	impressions_median	followers_min	followers_max	followers_mean	followers_median
0	0	0	0	0.0	0	0	371	7.6	4	0	164	49.8	34
0	1	1	2	1.3	1	0	537	27.8	16	0	164	59.5	43
0	2	3	7	3.9	3	2	653	113.9	55	4	164	71.5	67
0	3	9	55	23.9	18	0	3234	1154.6	679	43	155	107.1	106
1	0	0	0	0.0	0	0	253	9.2	6	165	938	447.3	409
1	1	1	2	1.3	1	0	383	26.1	17	165	939	468.3	407
1	2	3	8	4.1	4	4	1070	89.5	52	165	938	512.2	470
1	3	9	224	38.8	17	71	14126	1358.8	338	165	904	510.1	487
2	0	0	0	0.0	0	0	892	14.7	8	941	4209	2054.0	1903
2	1	1	2	1.3	1	1	744	35.9	19	947	4224	2092.5	1714.5
2	2	3	8	4.4	4	1	873	98.5	61	947	4207	2168.8	1968
2	3	9	470	30.3	16	78	80477	2397.1	327.5	1092	4216	2291.0	2220.5
3	0	0	0	0.0	0	0	2663	45.8	15	4273	4325387	35211.8	7329
3	1	1	2	1.5	1	3	5932	136.6	41	4269	4325387	122351.0	9922
3	2	3	8	4.8	4	13	12986	808.6	387	4317	4715644	384054.8	13053
3	3	9	56416	360.4	45	46	2310333	10846.3	2325	4415	4325387	224579.3	34101

Table 2: Clustering - Climate High Credibility

followers_cluster	engagement_cluster	engagement_min	engagement_max	engagement_mean	engagement_median	impressions_min	impressions_max	impressions_mean	impressions_median	followers_min	followers_max	followers_mean	followers_median
0	0	0	0	0.0	0	0	972	12.1	5	0	164	51.7	34
0	1	1	13	1.3	1	1	1927	31.9	19	0	164	63.2	50
0	2	3	12	1.2	1	0	2108	125.1	60	1	163	79.0	61.5
0	3	0	708	37.4	4	0	7173	798.5	108.5	0	163	79.0	61.5
1	0	0	0	0.0	24	0	454	17	0	0	938	466.4	433
1	1	1	1	1.3	0	0	1742	49.6	28	165	938	471.6	436
1	2	2	2	1.3	1	0	1742	49.6	28	165	938	471.6	436
1	3	8	8	4.3	4	5	11815	180.5	93	165	939	490.4	450
1	0	1170	1170	31.5	16	18	163759	2606.8	532.5	165	938	540.4	519
2	0	0	0	0.0	0	0	1177	29.5	15	940	4244	2052.0	1856
2	1	1	1	1.4	1	1	8648	76.8	45	940	4252	2148.4	1968
2	2	8	8	4.6	4	2	18673	234.1	136	942	4249	2211.1	2076
2	3	2123	2123	35.6	17	0	123313	2049.8	673	944	4244	2385.8	2295
3	0	0	0	0.0	0	0	4095	132.6	43	4253	3628769	29154.0	10382.5
3	1	1	1	1.4	1	0	5762	334.1	131	4253	8703458	76854.7	11508.5
3	2	8	8	4.9	5	0	14990	811.2	365	4254	14684441	236395.3	15505
3	3	7964	7964	79.7	25	16	1277216	7511.0	2019	4253	54998517	530118.4	34637

Table 3: Clustering - Covid Low Credibility

followers_cluster	engagement_cluster	engagement_min	engagement_max	engagement_mean	engagement_median	impressions_min	impressions_max	impressions_mean	impressions_median	followers_min	followers_max	followers_mean	followers_median
0	0	0	0	0	0	0	1313	6.5	3	0	93	30.5	24
0	1	1	1	1	1	0	785	23.6	12	0	93	34.4	28
0	2	5	2.6	2	2	1	2778	68.5	32	0	93	40.2	36
0	3	202	18.5	9	9	6	14797	601.1	240	0	93	46.7	47
1	0	0	0.0	0	0	0	703	8.9	4	84	526	247.1	216
1	1	1	1.0	1	1	0	921	24.7	14	94	526	263.5	238
1	2	5	2.7	2	2	1	6863	59.6	30	94	526	271.5	245
1	3	1284	21.1	9	9	11	34327	579.6	189	94	523	280.8	264
2	0	0	0.0	0	0	0	1337	12.1	7	527	2409	1144.9	989
2	1	1	1.0	1	1	0	2023	29.4	17	527	2409	1186.2	1046.5
2	2	5	2.7	2	2	0	774	23.6	12	527	2409	1186.2	1046.5
2	3	2068	25.2	10	10	16	27154	738.8	331	527	2408	1336.8	1284
3	0	0	0.0	0	0	0	2748	23.7	10	2410	4745856	11189.5	3715
3	1	1	1.0	1	1	0	4062	67.0	25	2415	4325447	24882.7	4318
3	2	5	3.0	3	3	2	10799	538.2	162	2411	4745856	86586.3	5349
3	3	104650	410.6	29	29	0	3624695	14448.5	1935	2412	11252986	198610.1	20109

Table 4: Clustering - Covid High Credibility

followers_cluster	engagement_cluster	engagement_min	engagement_max	engagement_mean	engagement_median	impressions_min	impressions_max	impressions_mean	impressions_median	followers_min	followers_max	followers_mean	followers_median
0	0	0	0	0	0	0	2982	13.4	6	0	93	32.7	26
0	1	1	1	1	1	0	807	33.2	17	0	93	37.1	33
0	2	2	2.7	2.7	2	0	687	85.1	38	0	93	39.3	36
0	3	0	16.7	0.0	0	0	19748	138.2	270	0	93	47.2	36
1	1	1	1.0	1.0	1	0	107	17.1	20	94	536	250.2	235
1	2	1	2.8	2.8	2	0	10188	30.0	20	94	536	269.6	242
1	3	5	20.6	20.6	10	0	7717	93.0	44	94	536	270.9	254
2	0	0	0.0	0.0	0	0	136671	1039.5	292	94	536	284.2	263
2	1	1	1.0	1.0	1	0	5875	22.6	13	527	2408	1221.9	1135
2	2	1	1.0	1.0	1	0	3797	46.1	25	527	2409	1217.6	1112
2	3	5	2.9	2.9	3	0	9747	108.0	55	527	2409	1249.9	1161
3	0	0	0.0	0.0	0	0	366959	1373.3	341	527	2409	1328.4	1276
3	1	1	1.0	1.0	1	0	8194	120.6	29	2410	8920800	29010.4	6114
3	2	2	3.1	3.1	3	0	7578	266.1	69	2410	14684264	67376.9	6892
3	3	5	173.3	173.3	22	0	11382	547.4	176	2410	20629935	156920.7	9275
3	6	109790	173.3	173.3	22	0	9488255	11384.8	1621	2410	18792639	382110.9	19597

2 Distribution of Engagement and Followers Data

To complement the information provided above, this section presents the log-log plots for the distribution of engagement and followers data.

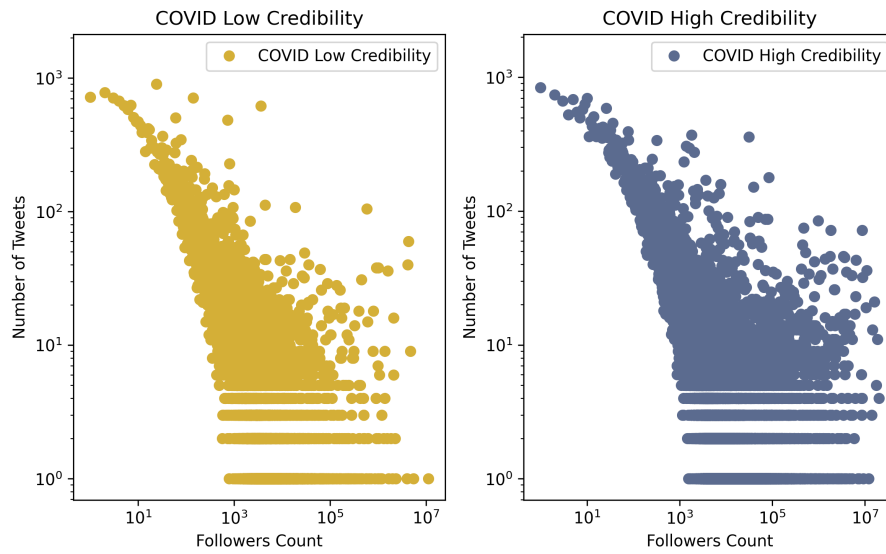


Figure 1: Log-Log Plot of Tweet Volume Versus Total Engagement for high-credibility and low-credibility COVID-Related Content

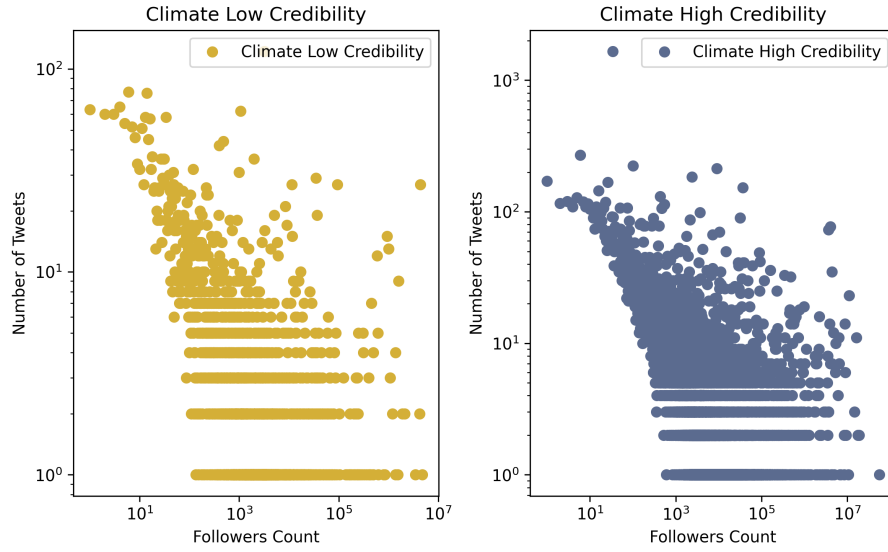


Figure 2: Log-Log Plot of Tweet Volume Versus Total Engagement for high-credibility and low-credibility Climate-Related Content

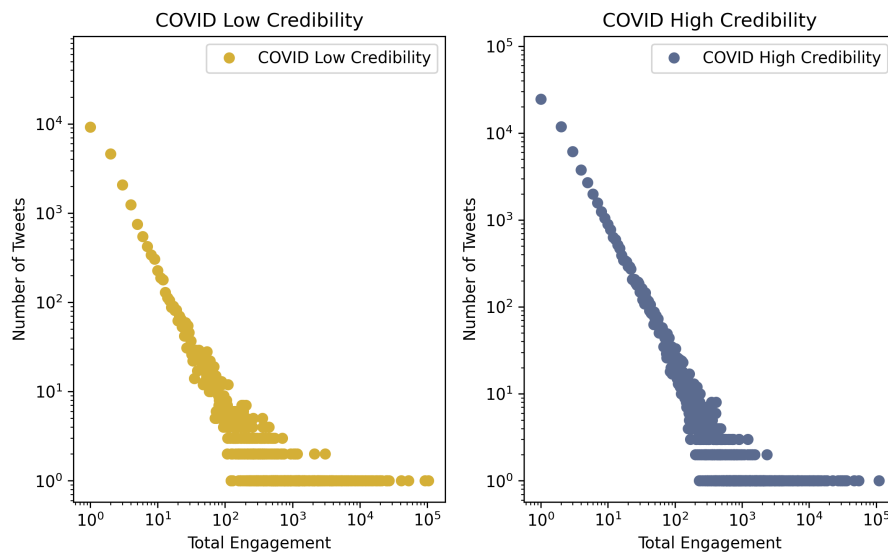


Figure 3: Log-Log Plot of Tweet Volume Versus Total User followers for high-credibility and low-credibility COVID-Related Content

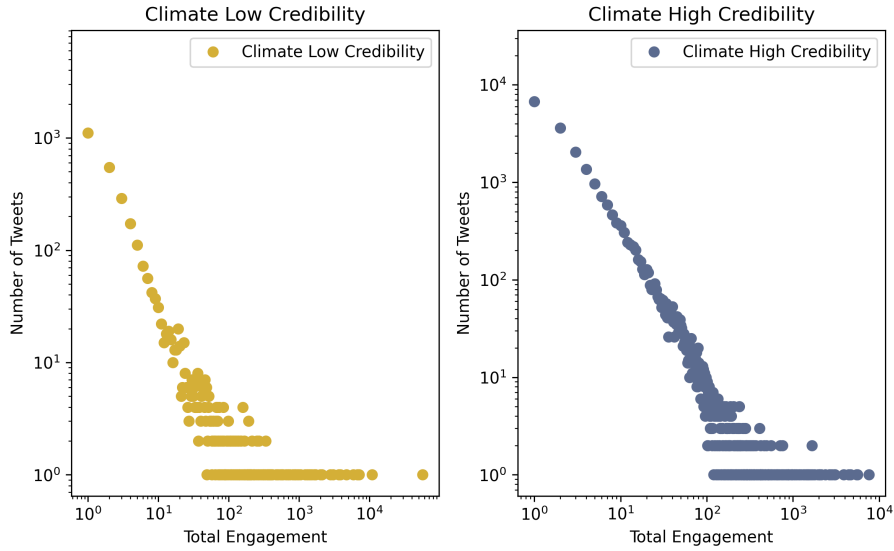


Figure 4: Log-Log Plot of Tweet Volume Versus Total User followers for high-credibility and low-credibility Climate-Related Content

3 Political Bias Annotation with the GPT-4 API

The political bias of URL domains was annotated using a custom-built zero-shot classifier, utilising the GPT-4 API. To achieve high replicability and to minimise the influence of randomness in responses, the temperature parameter was fixed at 0. The classifier was instructed through a system message: 'You are an assistant tasked with helping users identify the political bias of websites'. The main prompt used for bias assessment was formulated as follows: "Evaluate the political bias of the website 'domain' using the labels 'far-right', 'right', 'neutral', 'left', 'far-left'. If you lack information about the website, assign a score of -1". An example of the expected data format was provided to the model, though it was not shown any specific examples of the labelling task itself. This step was critical to ensure the clarity and uniformity of the data output.