Overlap Analysis of Online News Communities

Jaccard's Coefficient for Online News Communities

	TOI	HT	Hindu	IE	NIE	TGI	DH
TOI	1	**0.354**	**0.259**	**0.173**	0.015	0.002	0.001
HT		1	**0.349**	**0.267**	0.025	0.004	0.001
Hindu			1	**0.328**	0.037	0.006	0.002
IE				1	0.055	0.011	0.004
NIE					1	0.026	0.019
TGI						1	0.04
DH							1

Jaccard's Coefficient > 0.1 are highlighted.

Community Strength	
TOI	**11.4 Million**
HT	**6.56 Million**
Hindu	**5.01 Million**
IE	**2.96 Million**
NIE	**348 thousand**
TGI	**61.5 thousand**
DH	**25.3 thousand**

- HT, TOI, IE and Hindu (Set1) have larger community strength, comparable to each other, being in millions.
- NIE, TGI and DH (Set2) have relatively smaller community strength, comparable to each other, being in thousands.
- As such the jaccard's coefficient between overlap communities belonging to the two different sets(Set1 and Set2) is small. Thus, we analyse CDFs of overlap communities belonging to the same set.
- We can also see that communities belonging to Set1 have significant overlap, with high jaccard's coefficient (JC).

Policywise jaccard's coefficients for data (tweets) collected on the policy

Aadhar	TOI	HT	Hindu	IE	NIE	TGI	DH	Community Size
TOI	1.000							31368
HT	**0.494**	1.000						23107
Hindu	**0.421**	**0.452**	1.000					23733
IE	**0.425**	**0.482**	**0.478**	1.000				22518
NIE	0.123	0.151	0.149	0.164	1.000			5096
TGI	0.088	0.120	0.125	0.143	0.142	1.000		4736
DH	0.027	0.035	0.035	0.037	0.102	0.077	1.000	1183

GST	TOI	HT	Hindu	IE	NIE	TGI	DH	Community Size
TOI	1.000							21715
HT	**0.541**	1.000						16816
Hindu	**0.463**	**0.507**	1.000					16135
IE	**0.474**	**0.531**	**0.519**	1.000				15908
NIE	0.145	0.171	0.180	0.192	1.000			4109
TGI	0.095	0.120	0.131	0.145	0.165	1.000		3068
DH	0.033	0.041	0.044	0.046	0.121	0.093	1.000	943

Policywise jaccard's coefficients for data (tweets) collected on the policy

Demon	TOI	HT	Hindu	IE	NIE	TGI	DH	Community Size
TOI	1.000							37665
HT	**0.525**	1.000						29385
Hindu	**0.450**	**0.481**	1.000					28672
IE	**0.460**	**0.508**	**0.507**	1.000				28633
NIE	0.133	0.157	0.162	0.170	1.000			6464
TGI	0.097	0.124	0.134	0.147	0.157	1.000		5852
DH	0.029	0.037	0.038	0.039	0.106	0.084	1.000	1473

FP	TOI	HT	Hindu	IE	NIE	TGI	DH	Community Size
TOI	1.000							31720
HT	**0.541**	1.000						25318
Hindu	**0.473**	**0.499**	1.000					25911
IE	**0.490**	**0.532**	**0.524**	1.000				26358
NIE	0.162	0.186	0.185	0.191	1.000			6695
TGI	0.118	0.148	0.155	0.166	0.161	1.000		6180
DH	0.038	0.047	0.047	0.047	0.115	0.086	1.000	1626

We see the similar trend of communities belonging to Set1 having much more significant JC as compared to other overlap communities
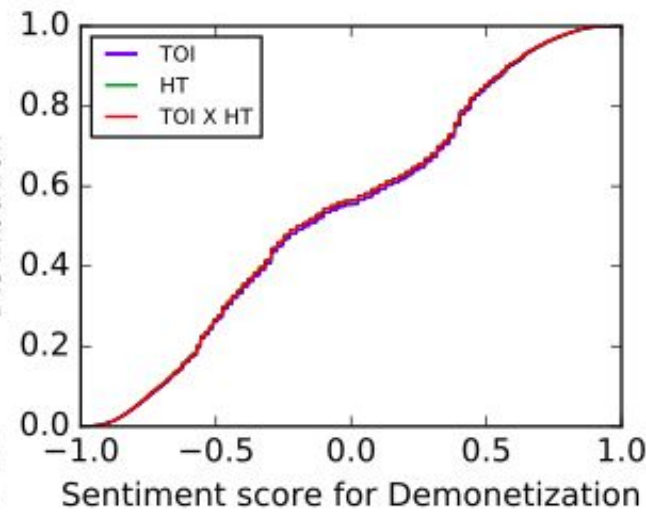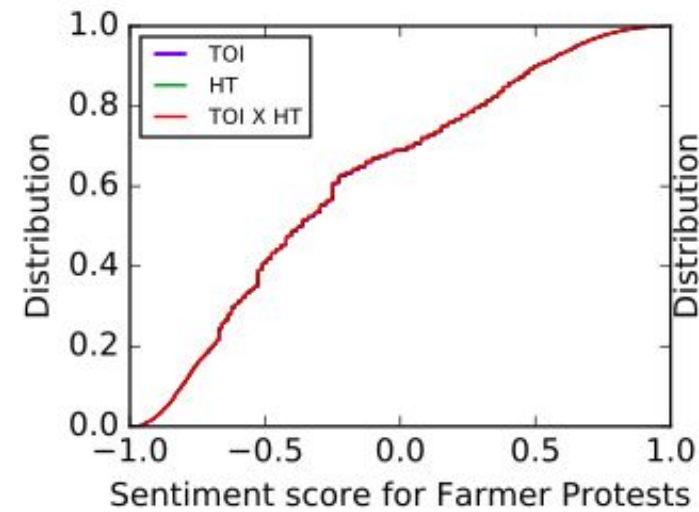
CDF plot for sentiment distribution for the overlap community between Hindustan Times and Times of India.

We see that CDF plots for HT,TOI and HTxTOI are quite close, agreeing with the large JC for these two newspapers.
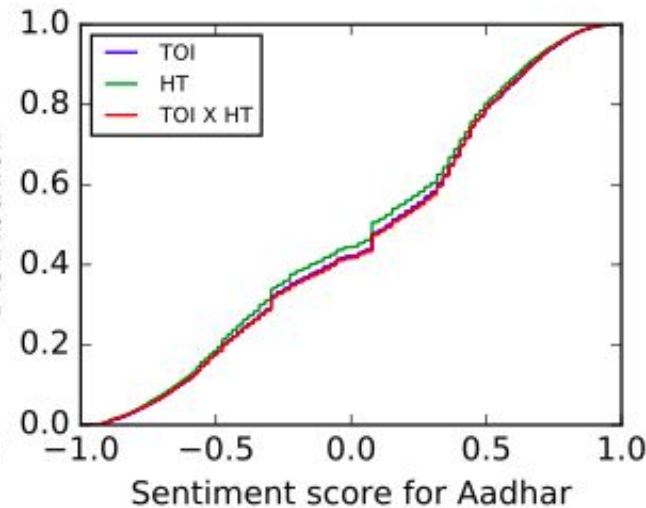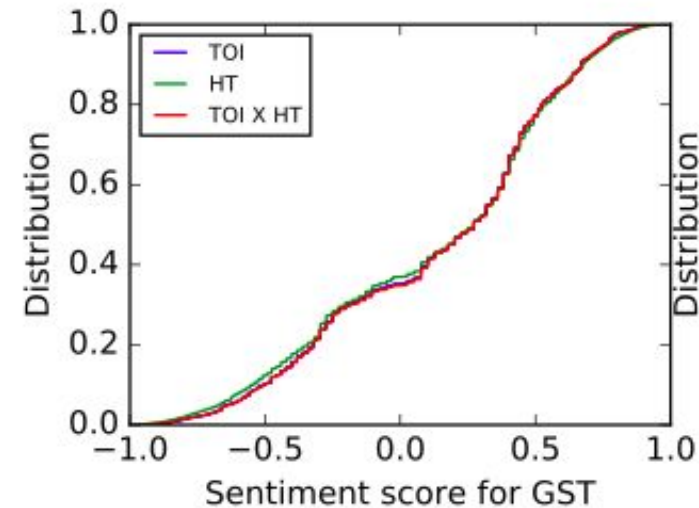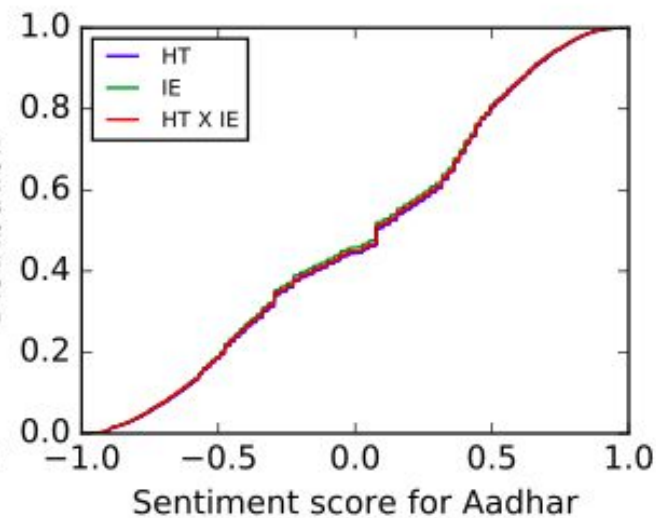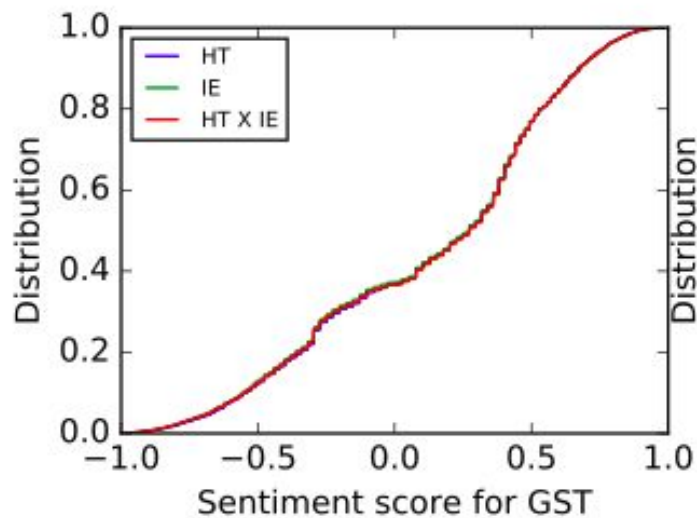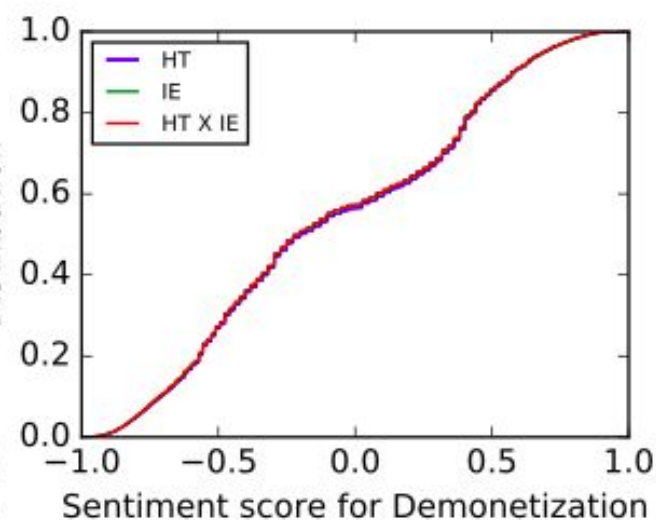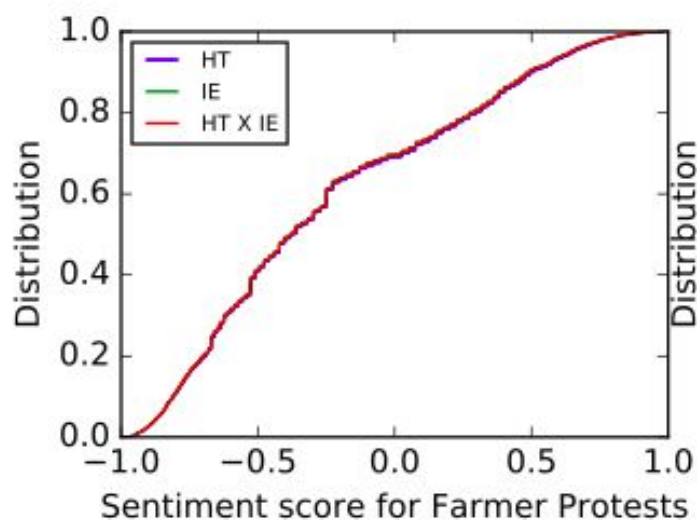
Similar trends are visible for other overlap communities with high JC values.
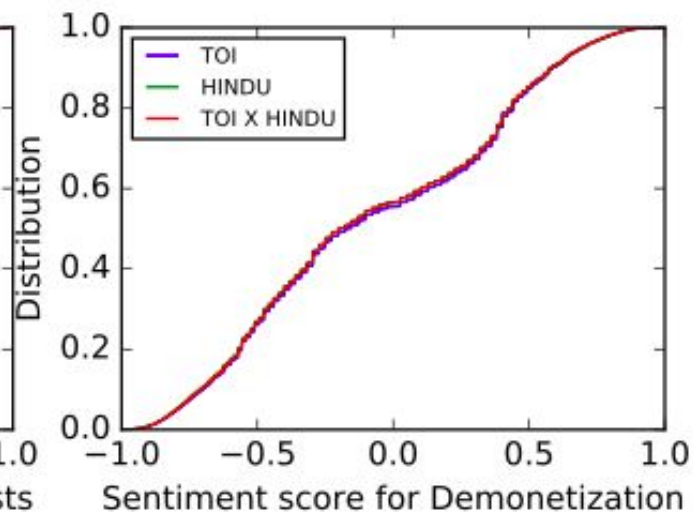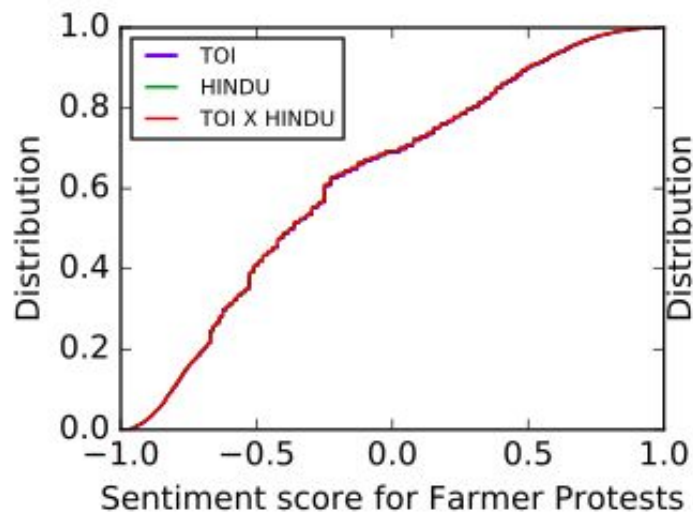
JC = **0.354**
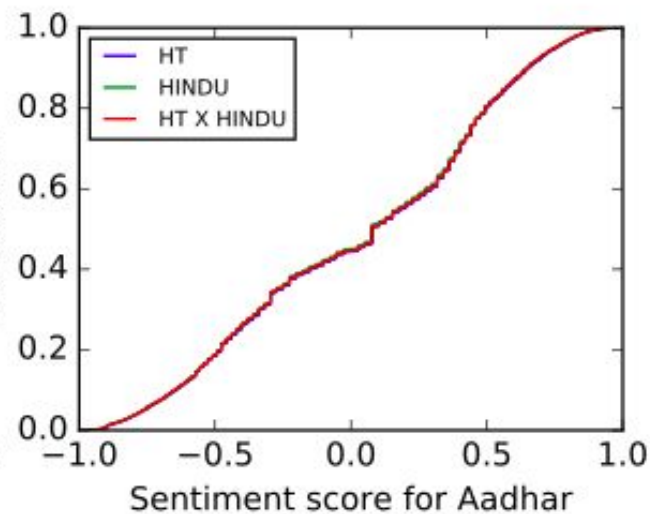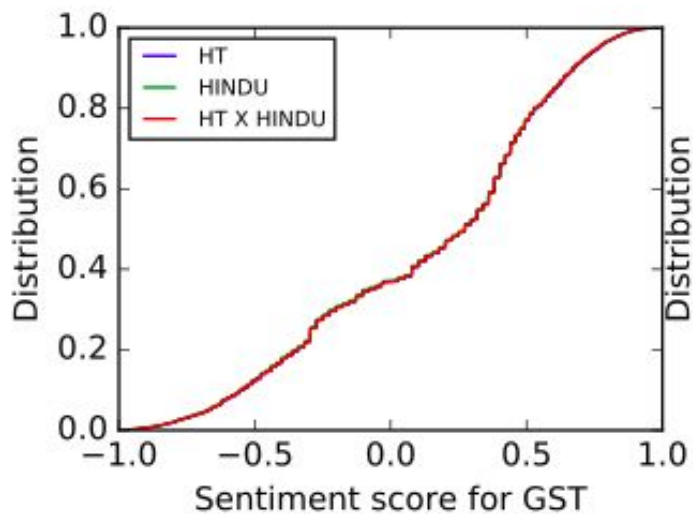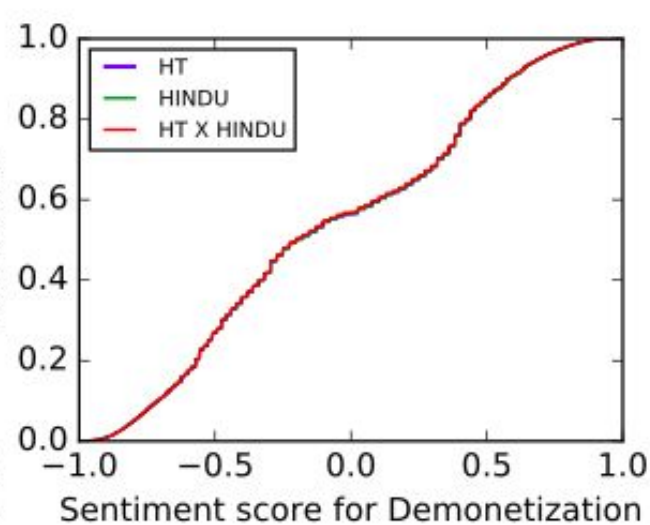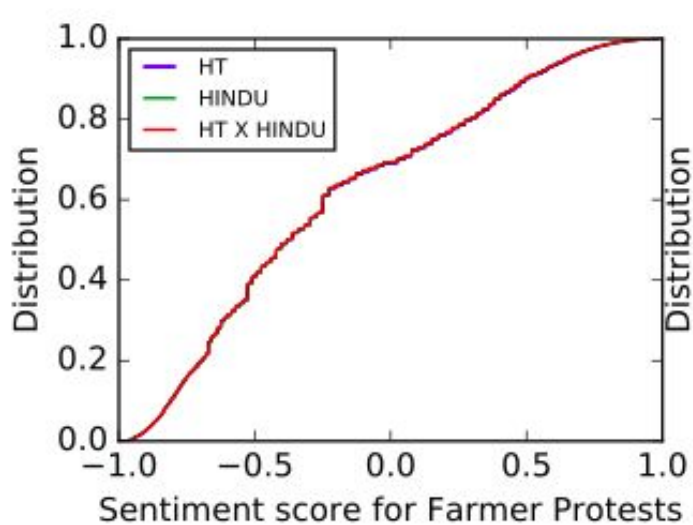
CDF plot for sentiment distribution for the overlap community between Hindustan Times and Hindu.
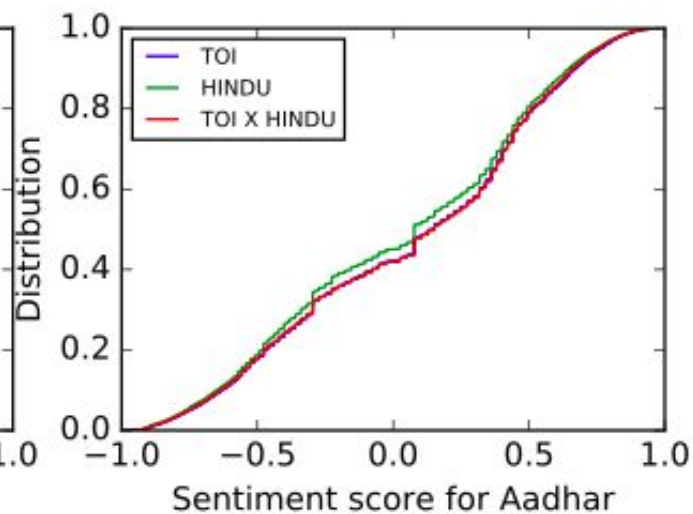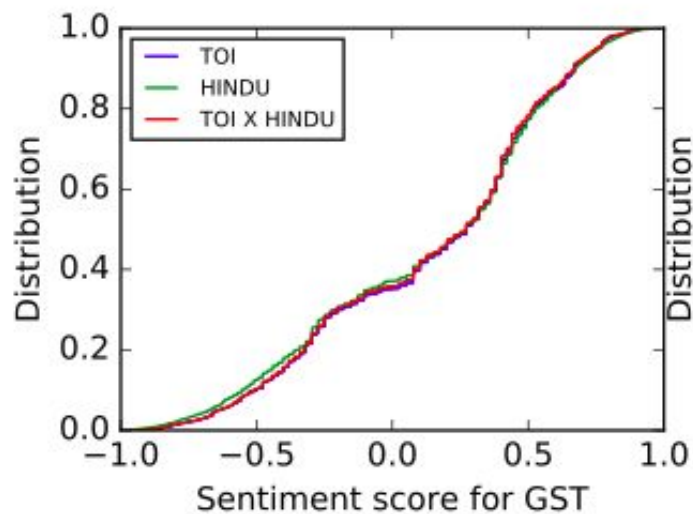
JC = **0.349**

CDF plot for sentiment distribution for the overlap community between Hindustan Times and IndianExpress.
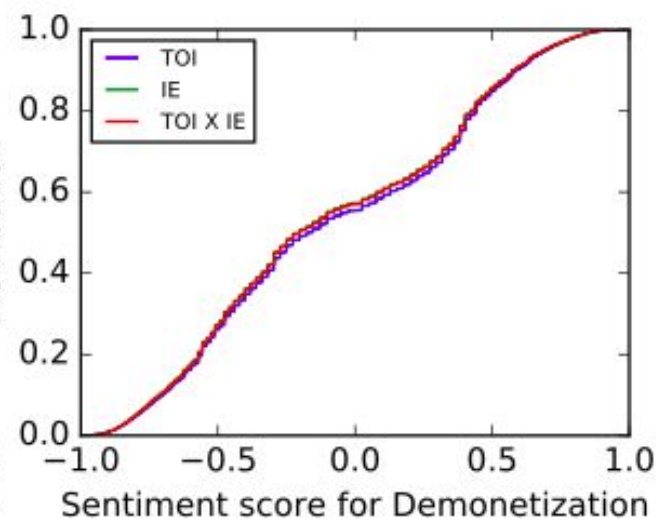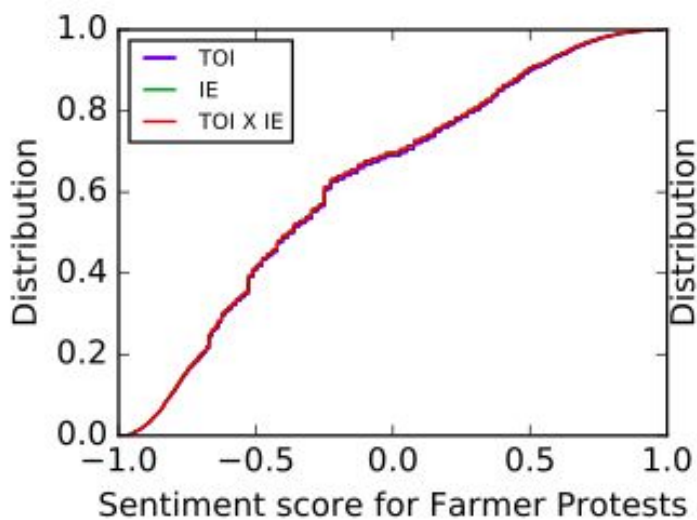
JC = **0.267**

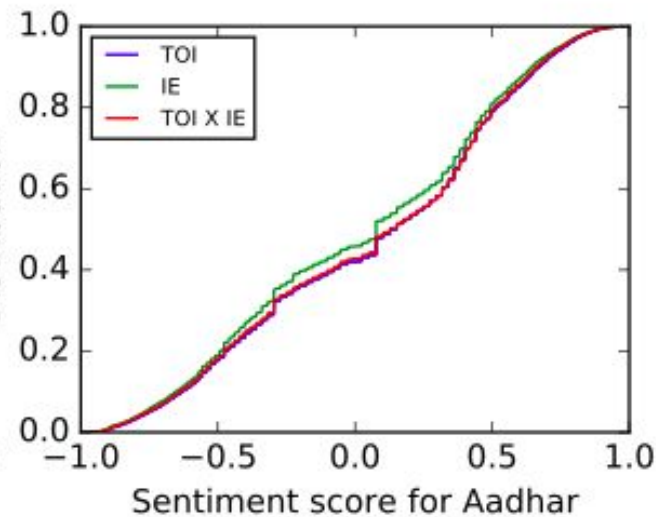CDF plot for sentiment distribution for the overlap community between Times of India and Hindu.
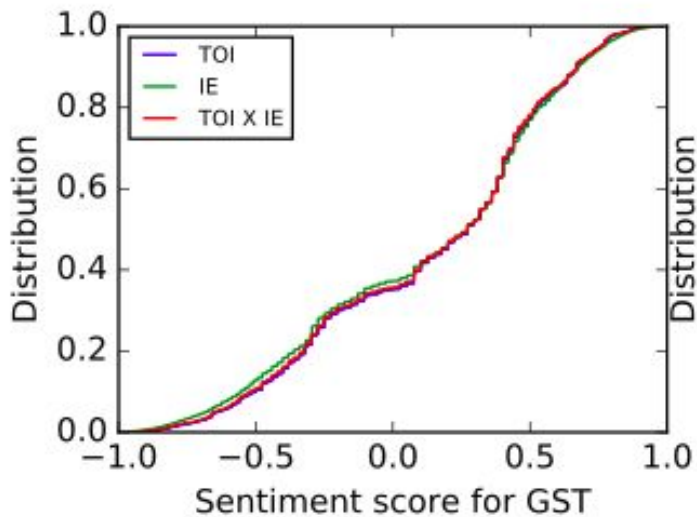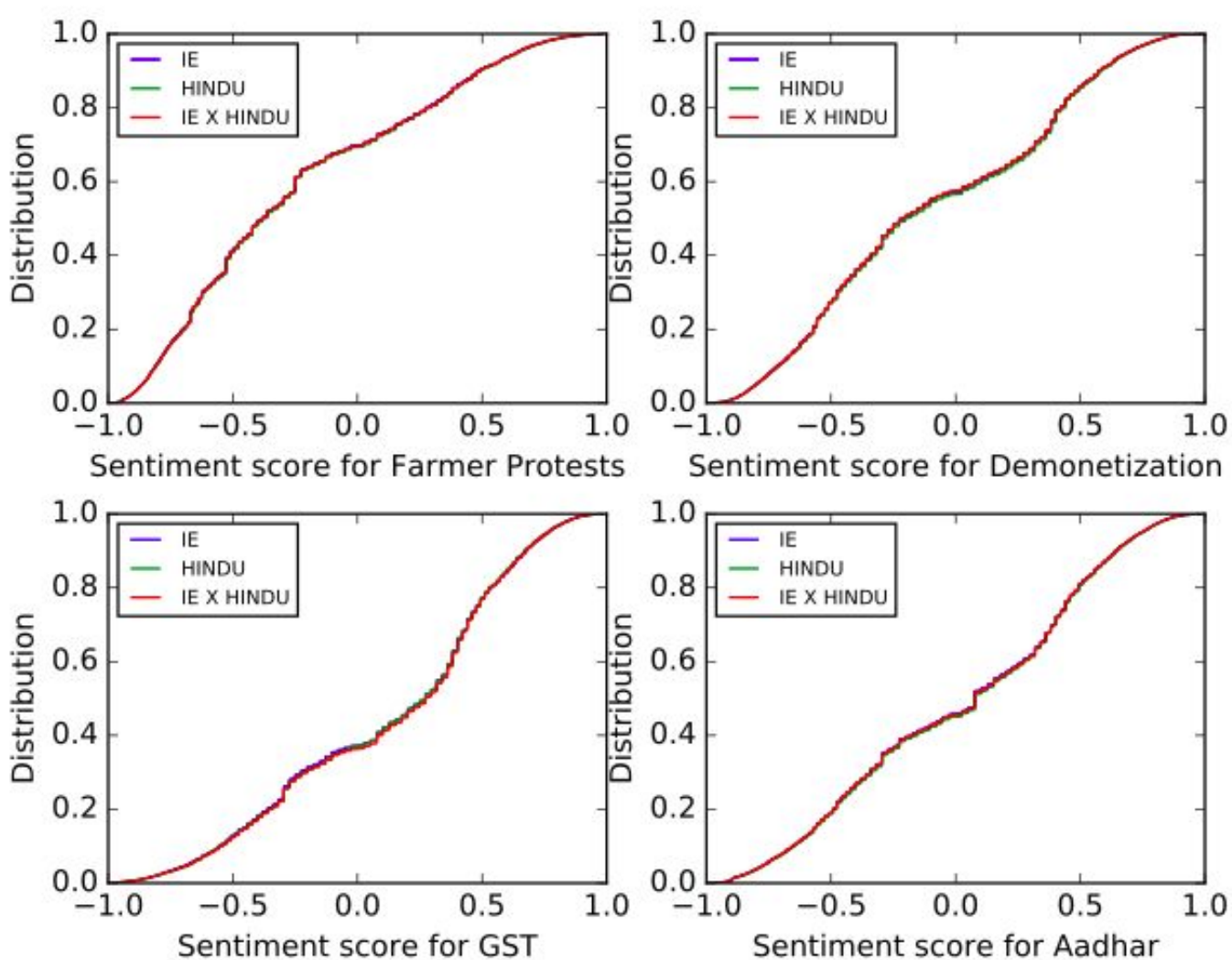
JC = **0.259**

CDF plot for sentiment distribution for the overlap community between Times of India and IndianExpress.

JC = **0.259**

CDF plot for sentiment distribution for the overlap community between Hindu and IndianExpress.
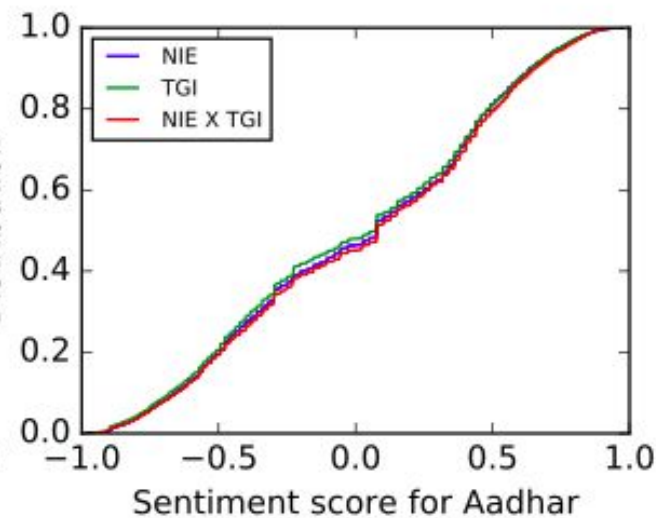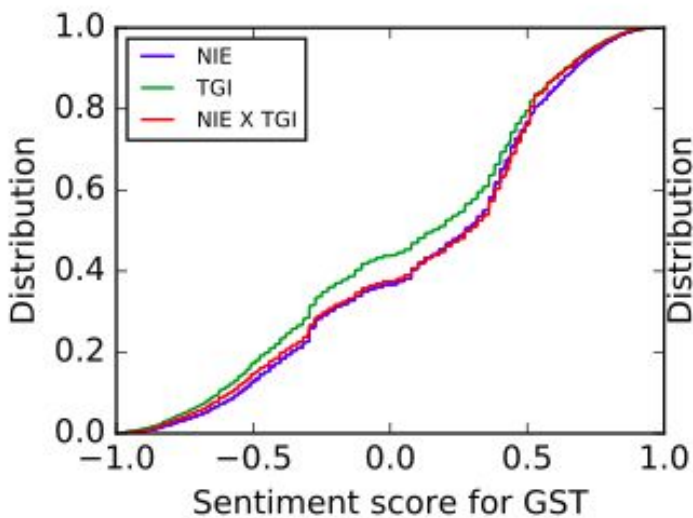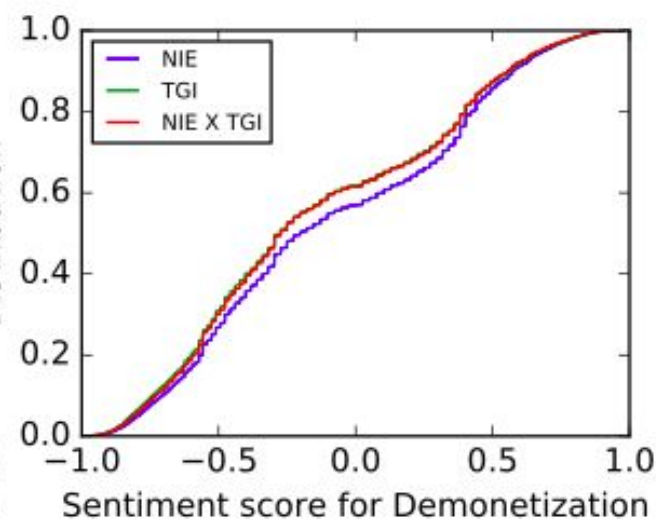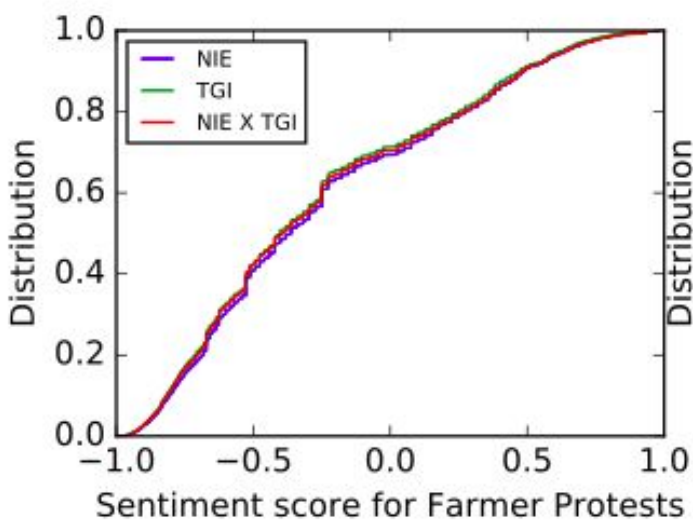
JC = **0.328**

CDF plot for sentiment distribution for the overlap community between NewIndianExpress and Telegraph India.

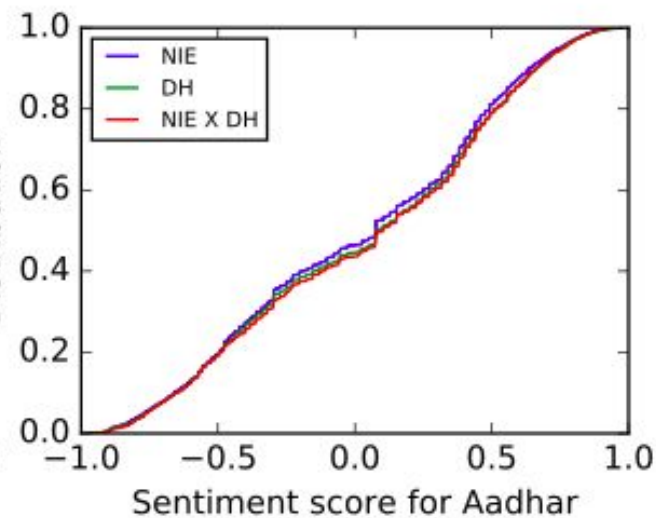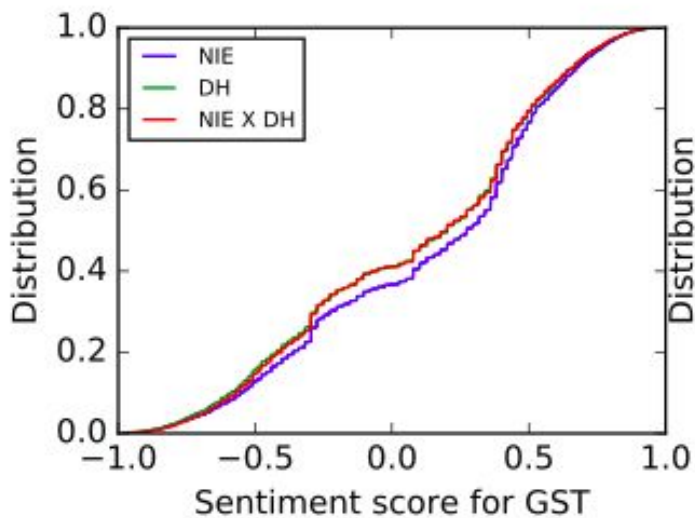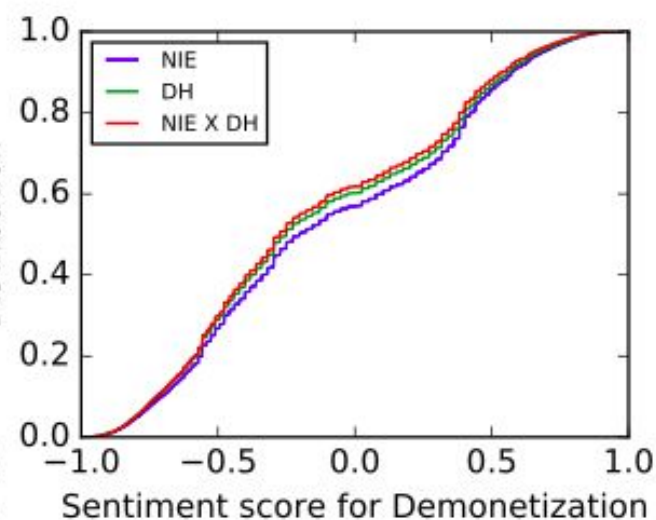We see that the CDF's for NIE, TGI and NIExTGI are not as close as for those with higher JC.

JC = **0.026**

CDF plot for sentiment distribution for the overlap community between NewIndianExpress and Deccan Herald.
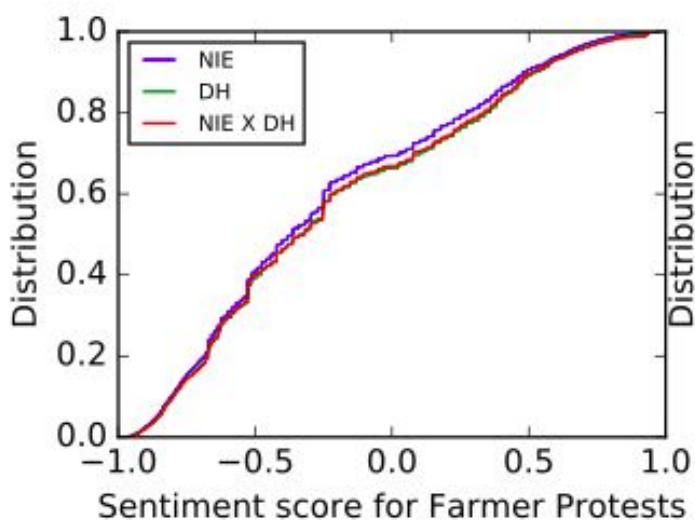
JC = **0.019**
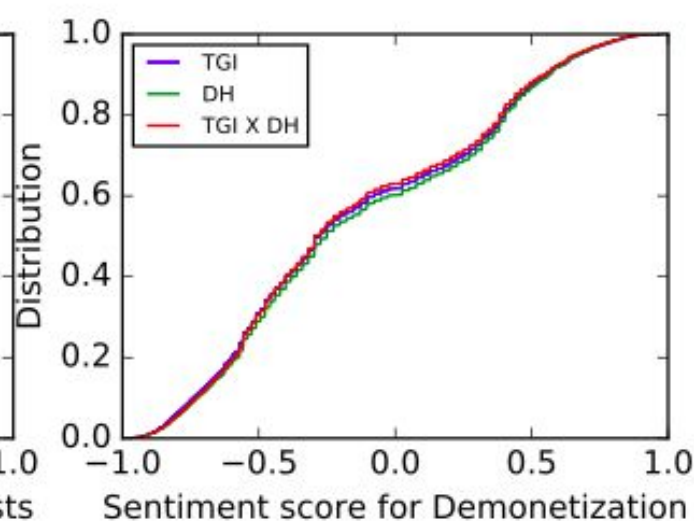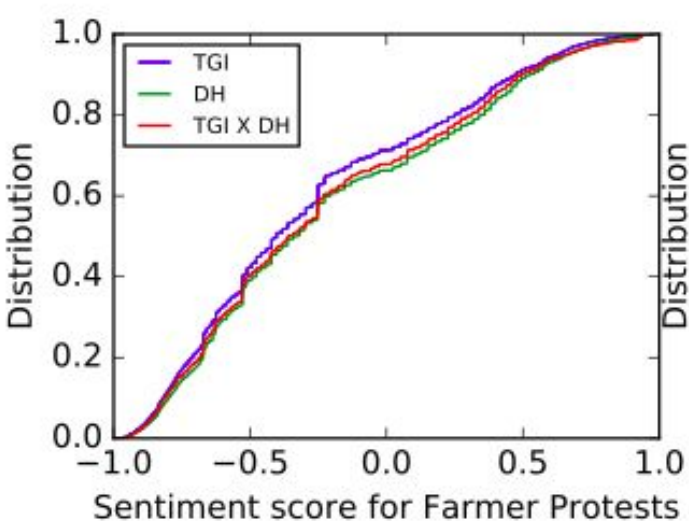
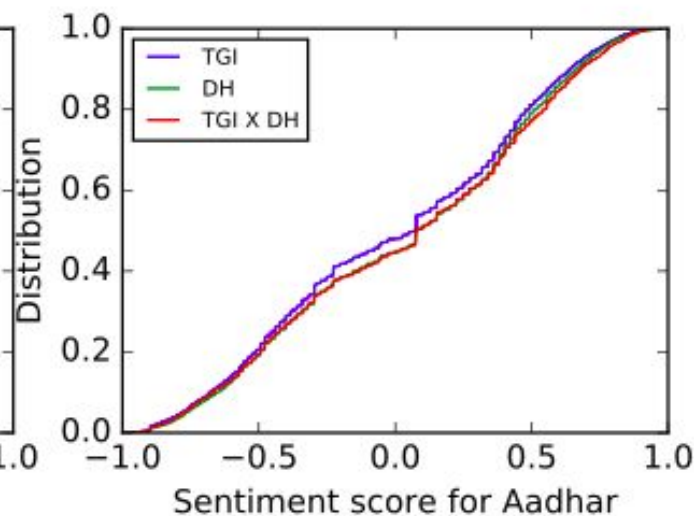CDF plot for sentiment distribution for the overlap community between Telegraph India and Deccan Herald.

JC = **0.04**
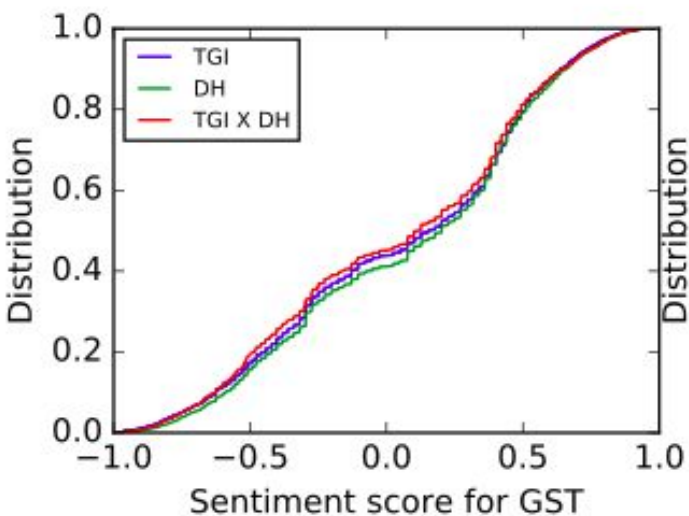
We see that the higher the extent of overlap, higher the similarity in sentiment expressed by the online news communities.

Overlap between online new communities and followers of political parties.

Table summarizing the overlap between the followers of newspapers with the two national parties, BJP and INC in terms of % newspaper followers following the party.

	BJP	INC	BJP-INC both	BJP ONLY	CONG ONLY
TOI	23.9	10.0	6.9	17.0	3.1
HT	26.9	14.0	9.9	17.0	4.2
NIE	32.7	21.5	15.2	17.6	6.4
TG	41.9	42.5	28.7	13.2	13.8
DH	39.9	33.7	24.4	15.6	9.4
IE	34.7	20.1	14.8	20.0	5.4
HINDU	29.6	16.2	11.9	17.7	4.3

- No. followers for **BJP - 10.4 million**
  No. followers for **INC - 4.7 million**

- It can be seen that at least 10% of follower base of newspapers also follow the political parties.

- The follower base of newspaper seems to follow BJP more than INC. This can be attributed to larger no. of followers of BJP (almost 2 ties) as compared to INC.

- Table for Jaccard's Coefficient

	BJP	INC	BJP-INC both	BJP ONLY	CONG ONLY
TOI	**0.1403**	0.0749	0.0574	**0.1130**	0.0274
HT	**0.1131**	0.0869	0.0724	0.0845	0.0333
NIE	0.0107	0.0148	0.0164	0.0079	0.0102
TG	0.0021	0.2109	0.0050	0.0009	0.0038
DH	0.0010	0.0018	0.0021	0.0005	0.0012
IE	0.0783	0.0794	0.0771	0.0567	0.0330
HINDU	0.1039	0.0886	0.0800	0.0750	0.0321

Table showing number of users in the overlap communities between newspaper followers and political parties.

		BJP	INC	BJP-INC both	BJP ONLY	CONG ONLY
	Total	10371874	4765818	2917478	7454396	1848340
TOI	11026375	2632738	1099989	756532	1876206	343457
HT	6299717	1694200	885066	622274	1071926	262792
NIE	347149	113645	74702	52635	61010	22067
TG	51885	21765	22047	14905	6860	7142
DH	24897	9938	8392	6064	3874	2328
IE	2742133	952445	552184	405355	547090	146829
HINDU	4842235	1432241	782379	574476	857765	207903