SAPIENZA
UNIVERSITÀ DI ROMA

Master's Degree in Data Science

MAPPING NEWS DIET AND POLARIZATION DYNAMICS DURING ELECTORAL CAMPAIGNS

Presented by Marco D'Ercole

Advisor: Walter Quattrociocchi

ANALYSIS QUESTION

QUESTION

How does affective polarization emerge by observing the discussion of political parties and newspapers on the social media platform?

METHOD

To answer this question, we will examine the level of sparsity of clusters to see whether a cluster structure emerges.

SAPIENZA
UNIVERSITÀ DI ROMA

MY DATASETS

DATASET PARTIES

Consisting of the posts published on Facebook by Italian political parties from 1 January to the day of the European elections (9 June 2024)

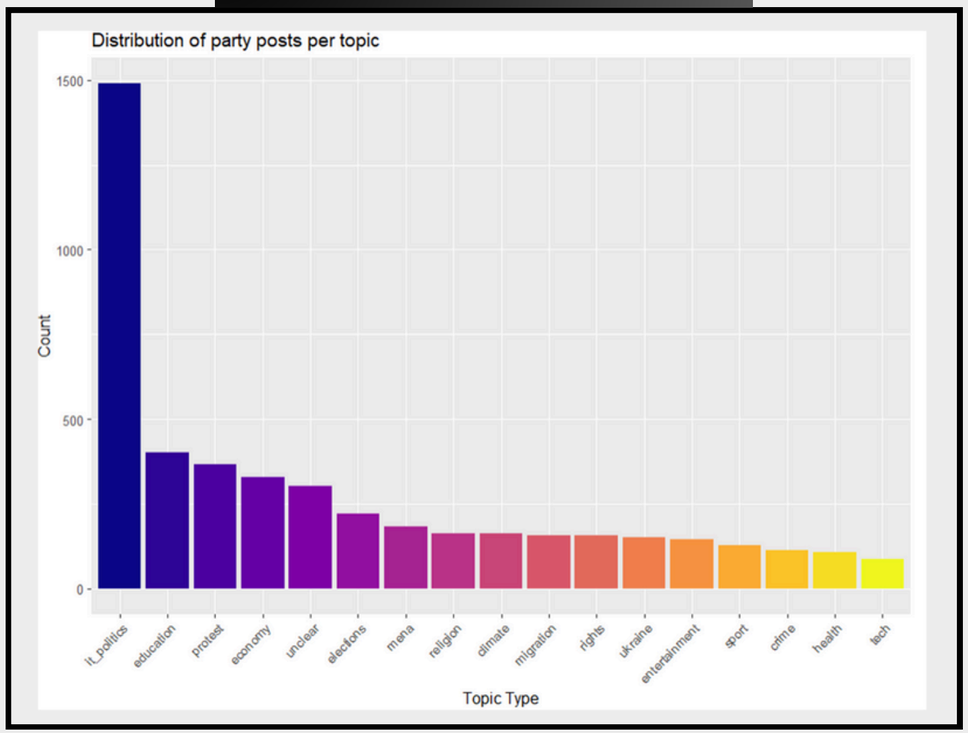
DATASET NEWS

Consisting of the posts published on Facebook by Italian newspeaper from 1 January to the day of the European elections (9 June 2024)

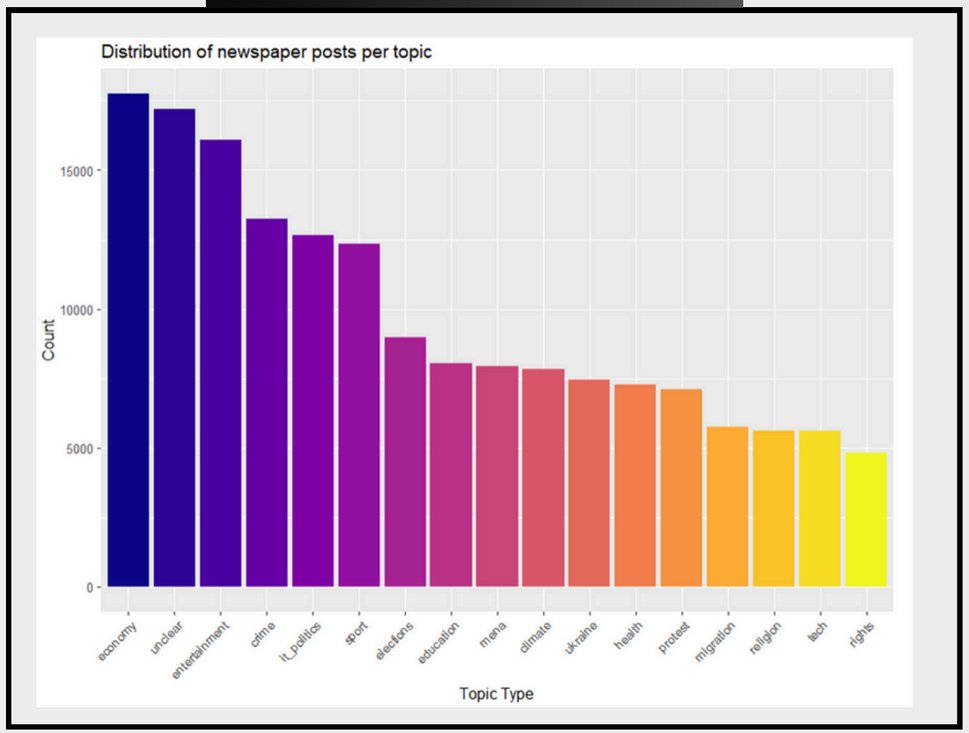
TOPICS DISTRIBUTIONS

SAPIENZA
UNIVERSITÀ DI ROMA

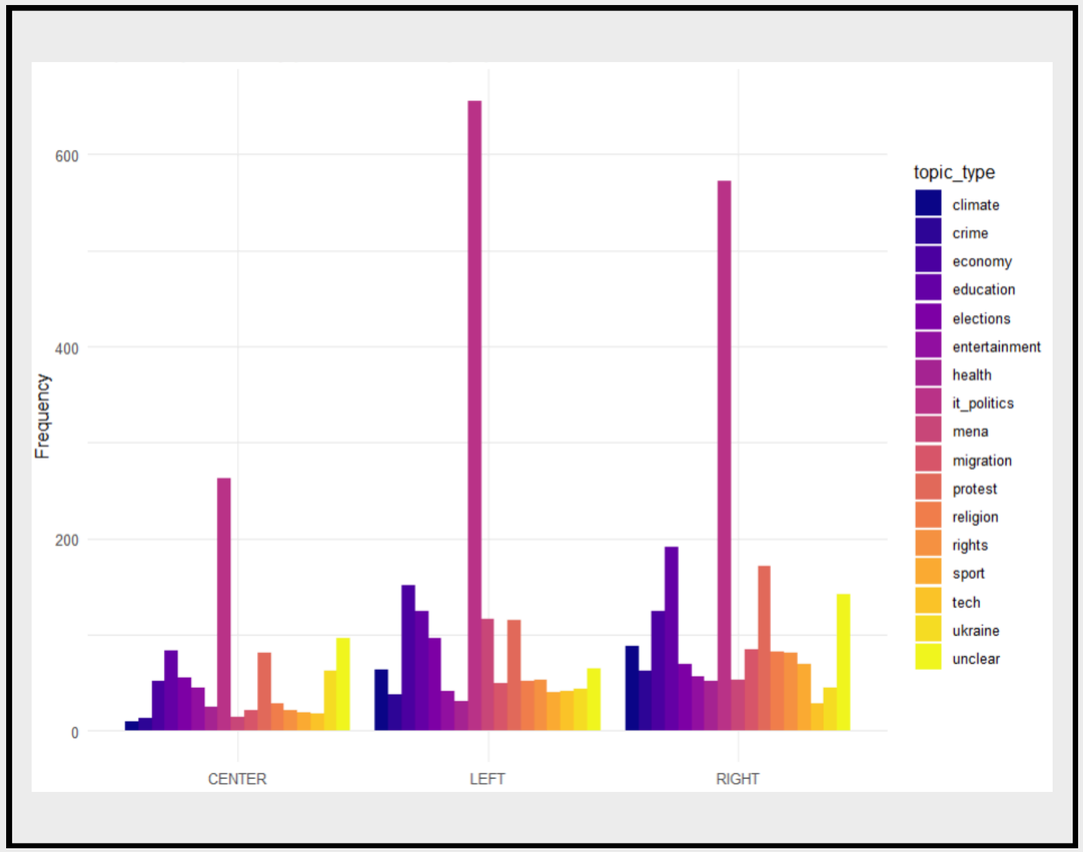
DATASET PARTIES



DATASET NEWS



POSTS AND POLITICAL TREND



Political parties

CENTER

It_politics, education, protest, **Ukraine,** economy**, elections, entertainment**

LEFT

It_politics, economy, education, **mena**, protest, **elections**, **climate**

RIGHT

It_politics, education, protest, economy, **climate**, **migration**, **religion**

THE MOST DISCUSSED TOPICS

Without considering it_politics the highest values are for:

- **Azione**: protest
- **Verdi**: economy
- **Forza Italia**: economy
- **Fratelli d'Italia**: economy
- **Italia Viva**: education
- **Matteo Salvini**: education
- **Movimento 5 Stelle**: economy
- **Partito Democratico**: education
- **Sinistra Italiana**: mena

Percentage of posts for each party according to the topic

topic_type	Azione	Europa Verde - Verdi	Forza Italia	Fratelli d'Italia	Italia Viva	Matteo Salvini	MoVimento 5 Stelle	Partito Democratico	Sinistra Italiana
climate	1.18	7.12	0.80	1.90	1.03	6.76	3.18	2.82	1.89
crime	0.71	2.54	0.00	0.63	2.05	5.30	1.59	2.54	2.08
economy	7.08	9.67	15.66	8.07	4.52	3.20	9.94	5.37	8.49
education	6.37	6.87	3.21	6.33	11.70	13.15	8.75	6.21	6.04
elections	5.42	5.85	7.63	3.16	6.78	2.83	5.77	5.08	4.91
entertainment	6.13	1.53	1.20	0.95	3.90	4.38	2.78	3.67	1.70
health	3.07	2.04	2.41	1.58	2.46	3.29	0.60	1.69	2.64
it_politics	13.44	29.77	33.73	55.70	42.30	12.42	35.79	48.31	35.28
mena	1.42	6.87	2.41	2.37	1.64	2.92	1.99	3.11	12.83
migration	2.36	3.56	5.22	3.80	2.26	4.38	1.79	3.39	2.83
protest	14.62	7.12	5.22	5.06	3.90	11.60	4.37	4.80	9.06
religion	1.42	2.04	4.02	1.58	4.52	5.75	5.96	1.13	1.89
rights	3.54	3.56	4.82	1.58	1.44	5.39	1.99	3.11	3.40
sport	0.94	2.29	4.42	1.58	3.08	4.38	3.18	2.82	0.94
tech	3.54	2.54	1.61	1.27	0.62	1.46	4.17	0.56	1.70
ukraine	9.20	3.05	3.61	0.95	4.93	2.74	1.99	3.11	2.08
unclear	19.58	3.56	4.02	3.48	2.87	10.05	6.16	2.26	2.26

SHAPIRO-WILK TEST

Are the contents of newspapers and parties related? To proceed with the study of the correlation it is necessary to verify the normality of the data.
two examples of results:

PARTY

```
$Azione

        Shapiro-Wilk normality test

data:  newX[, i]
W = 0.84495, p-value = 0.00899
```

NEWSPAPER

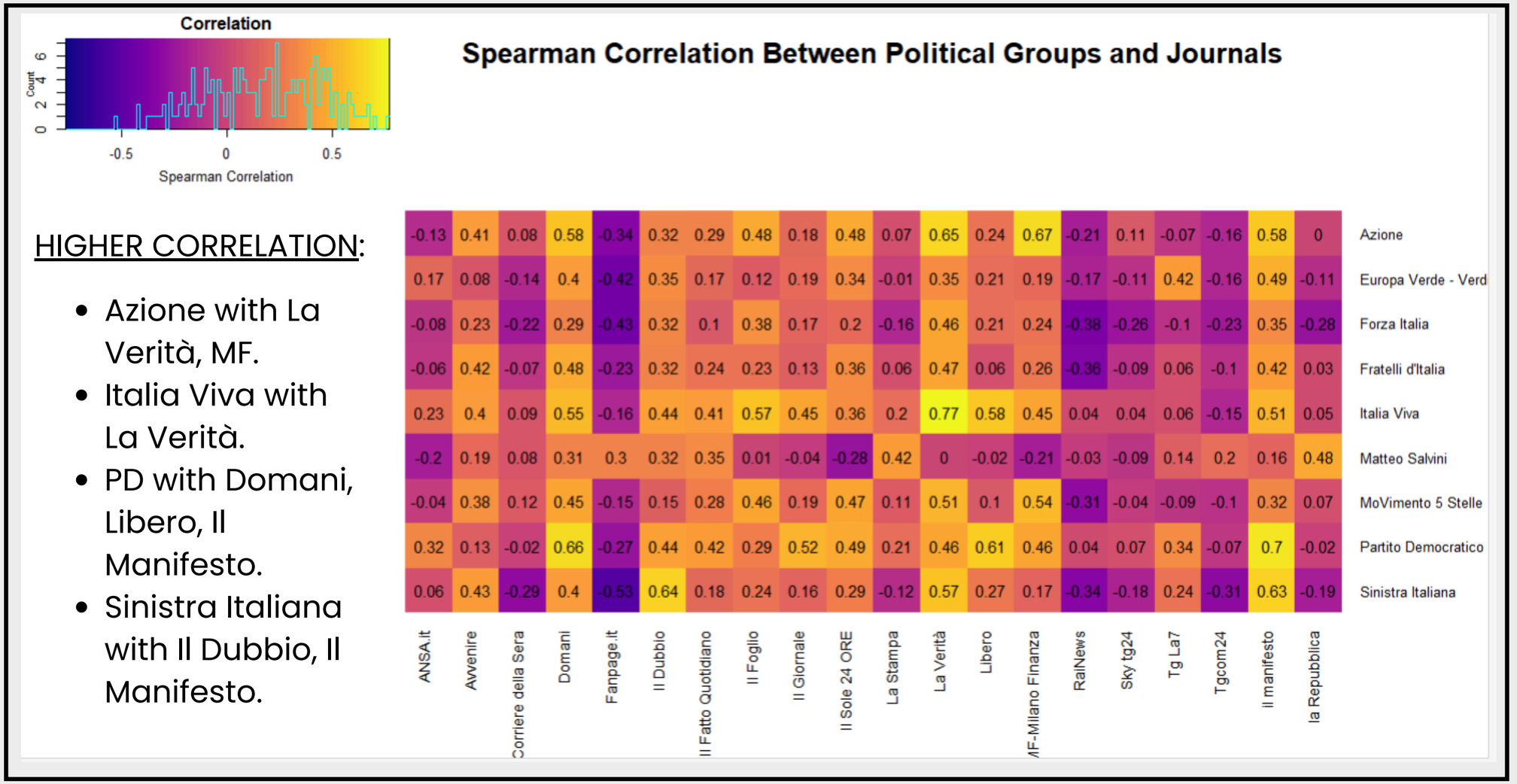
```
$`Corriere della Sera`

        Shapiro-Wilk normality test

data:  newX[, i]
W = 0.79217, p-value = 0.001588
```

The data do not follow a normal distribution ($p<0.05$) so we cannot use Pearson's correlation but will use Spearman's correlation.

SPEARMAN CORRELATION



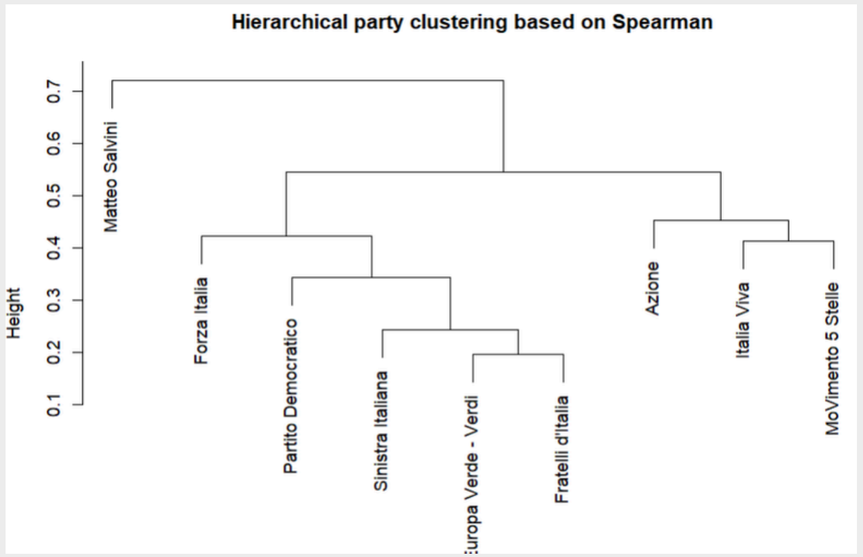
HIGHER CORRELATION:

- Azione with La Verità, MF.
- Italia Viva with La Verità.
- PD with Domani, Libero, Il Manifesto.
- Sinistra Italiana with Il Dubbio, Il Manifesto.

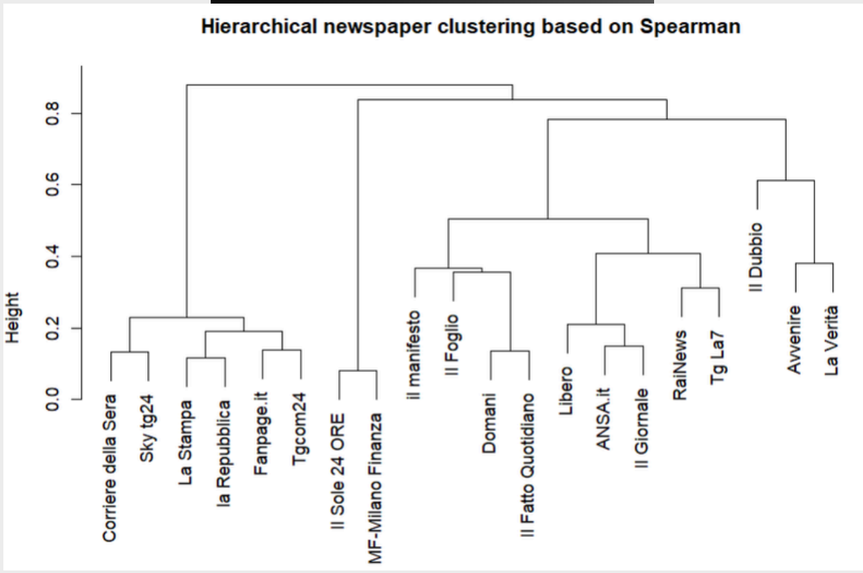
CLUSTERIZATION

Clustering is a process that allows us to group together parties and newspapers that possess the same characteristics, in this case talking about similar topics.

DATASET PARTIES

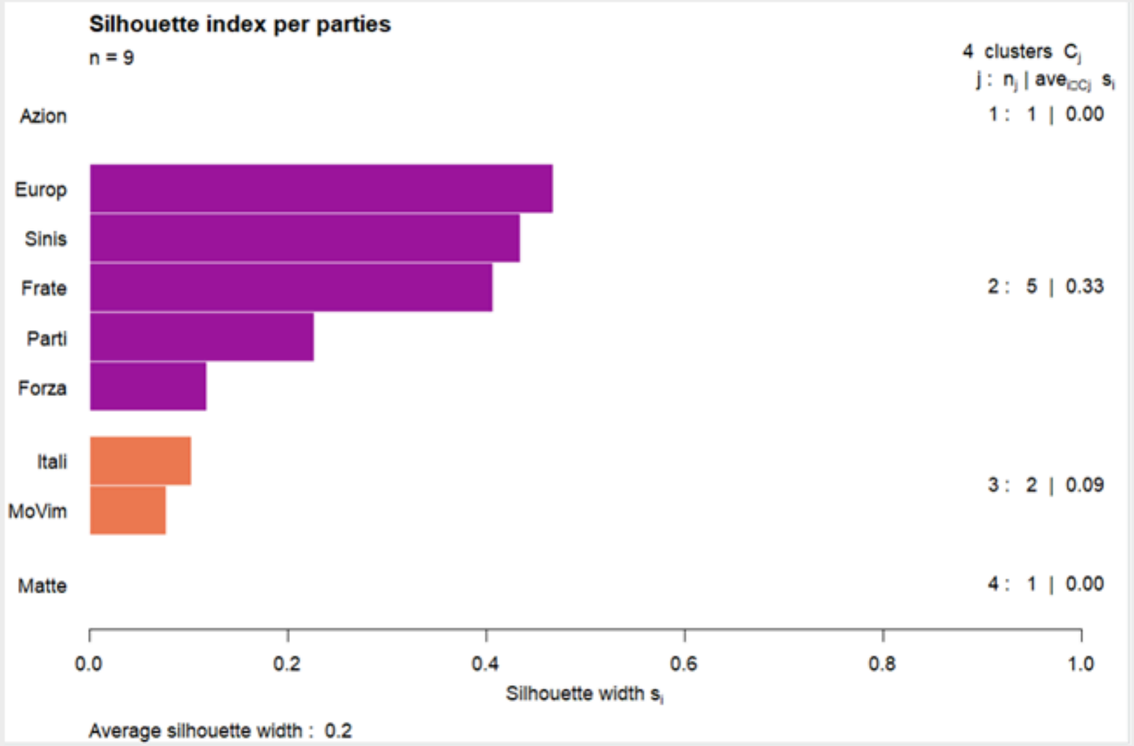


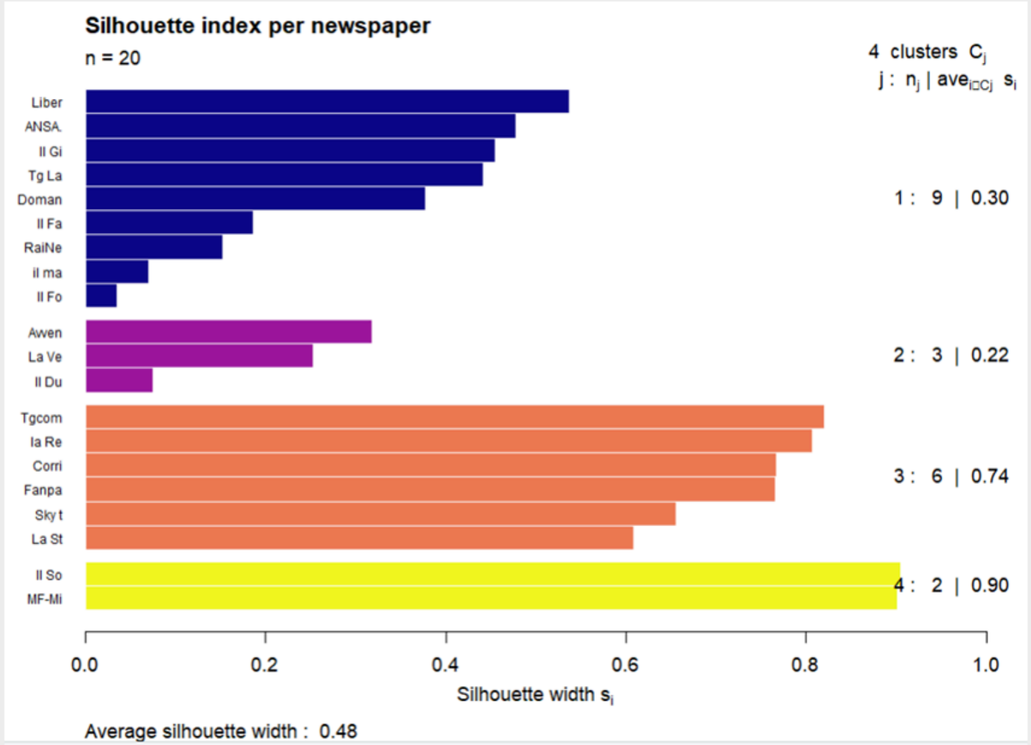
DATASET NEWS



I used the Spearman method for the correlation between parties/newspapers, then I applied hierarchical clustering with the method "Average".

HOW MANY CLUSTERS?





The silhouette index measures how well the data have been grouped into distinct clusters, through which the best subdivision can then be decided. This index combines two key factors: Cohesiveness and Separation. Through the silhouette index the choice on the number of clusters for the two datasets falls on k=4.

4 CLUSTERS!

SAPIENZA UNIVERSITÀ DI ROMA

DATASET PARTY

01
- Azione

02
- Europa Verde - Verdi
- Forza Italia
- Fratelli d'Italia
- Partito Democratico
- Sinistra Italiana

03
- Italia Viva
- Movimento 5 Stelle

04
- Matteo Salvini

DATASET NEWS

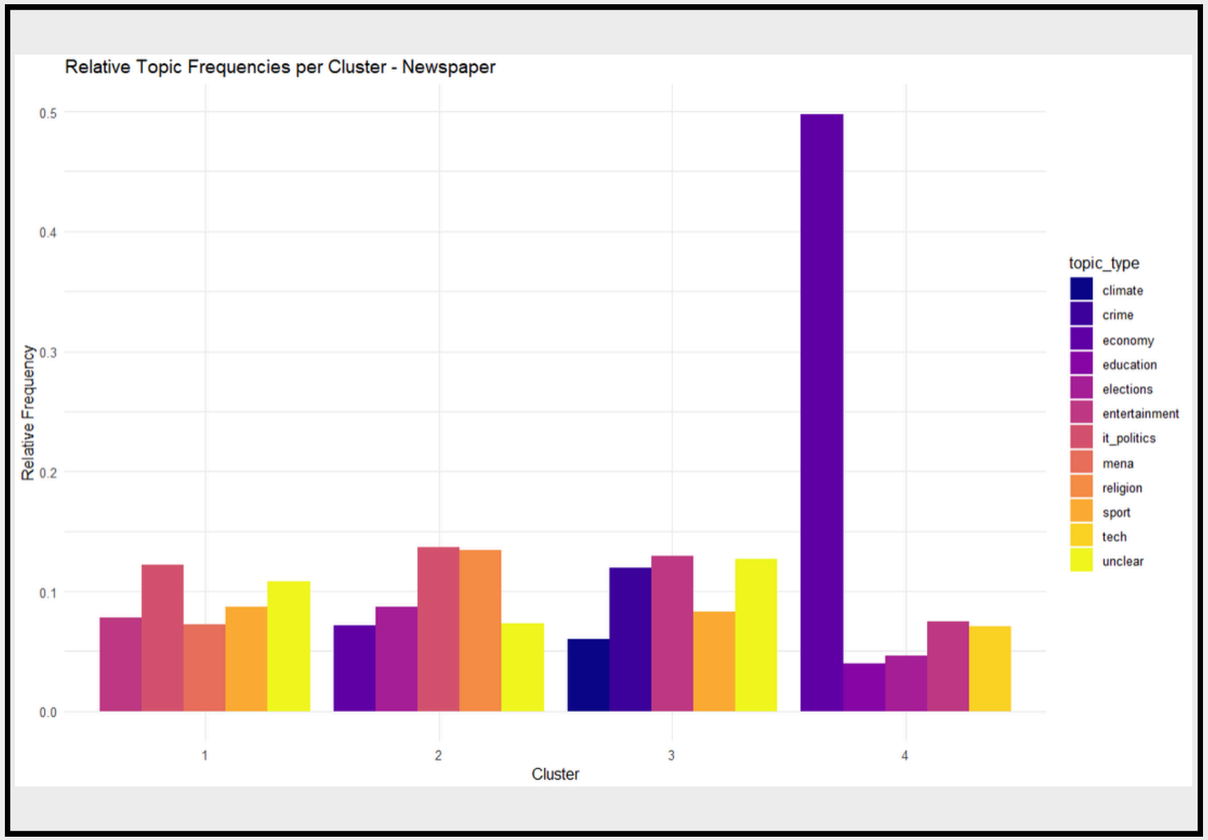
01
- Ansa.it
- Domani
- Il Fatto quotidiano
- Il Foglio
- Il Giornale
- Libero
- RaiNews
- TgLa7
- Il Manifesto

02
- Avvenire
- Il Dubbio
- La Verità

03
- Corriere della Sera
- Fanpage.it
- La Stampa
- Skytg24
- Tgcom24
- La Repubblica

04
- Il Sole24Ore
- MF - Milano Finanza

CLUSTERS FOR NEWS



CLUSTER 1 Ansa, Il Fatto Quotidiano, RaiNews...

It_politics, Sport, Entertainment, Mena.

CLUSTER 2 Avvenire, il dubbio, la verità

It_politics, Religion, Elections, Economy

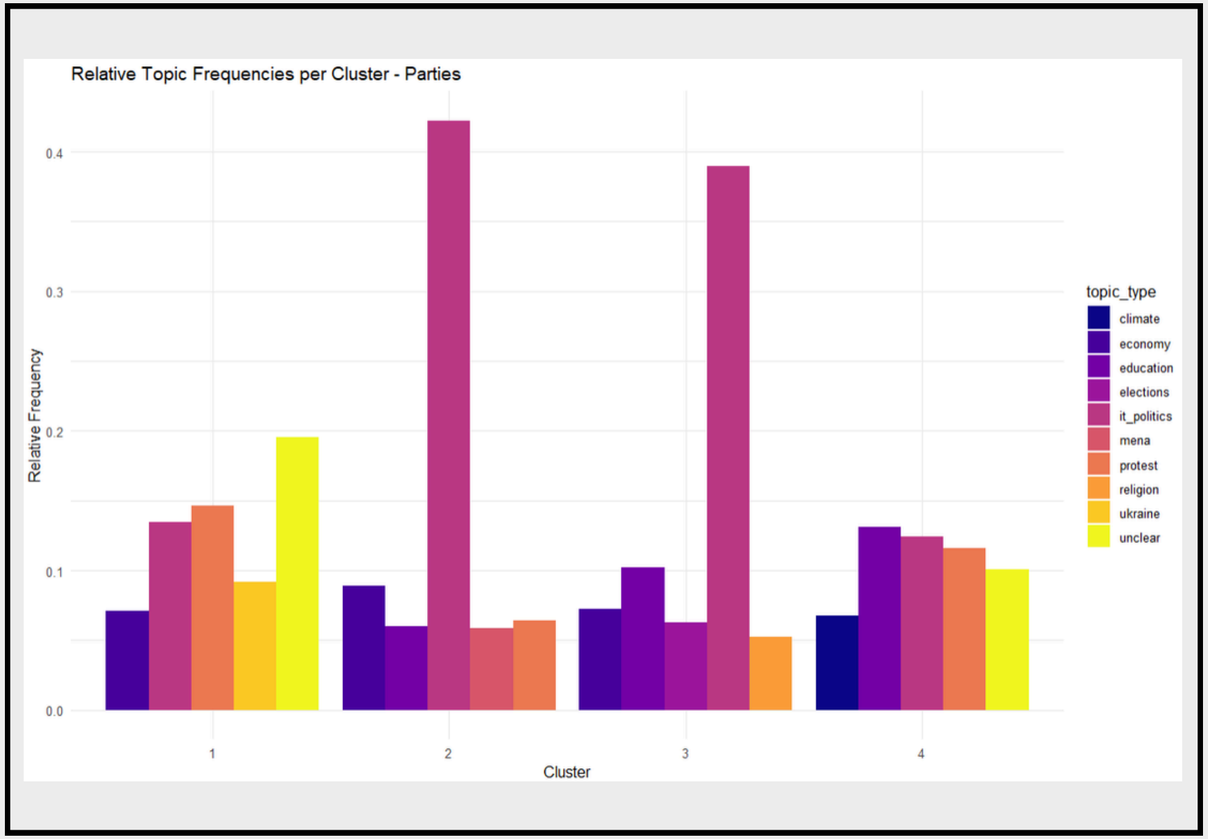
CLUSTER 3 Corriere della Sera, La Repubblica...

Entertainment, Crime, Sport, Climate

CLUSTER 4 Il Sole24Ore, Milano Finanza

Economy, Entertainment, Tech, Elections, Education

CLUSTERS FOR PARTIES



CLUSTER 1 Azione

Protest, It_politics, Ukraine, Economy

CLUSTER 2 Partito Democratico, Fratelli d'Italia...

It_politics, Economy, Protest, Mena, Education

CLUSTER 3 Italia Viva, Movimento 5 Stelle...

It_politics, Education, Economy, Elections, Religion

CLUSTER 4 Matteo Salvini

Education, It_politics, Protest, Climate

ENTROPY

THE QUESTION

How to study the bias of a party or a newspaper towards a particular topic?

THE ANSWER

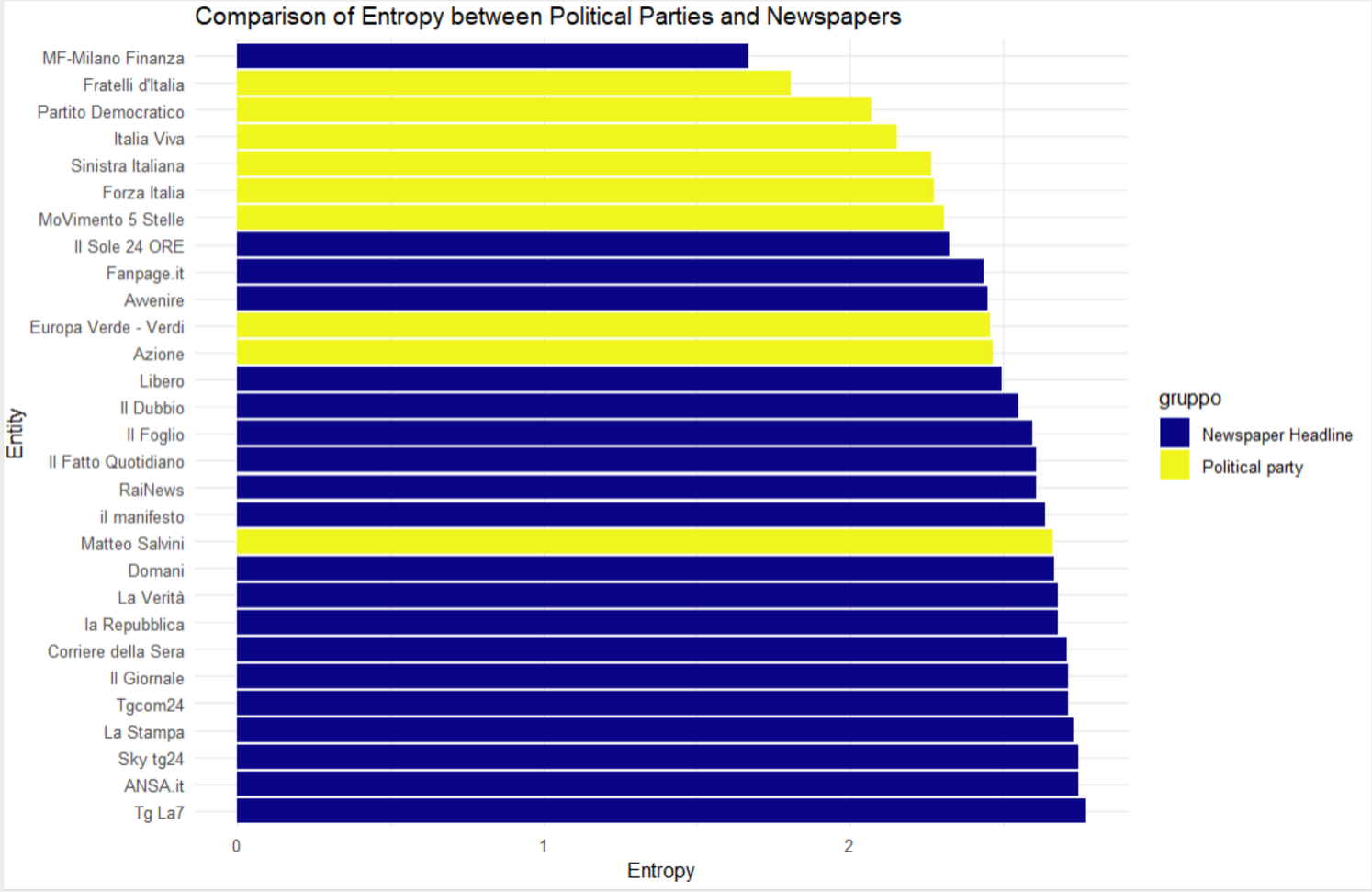
Entropy is a measure of the disorder or uncertainty associated with a probability distribution.
A distribution with higher entropy is more ‘dispersed’ or random, while one with lower entropy is more ‘concentrated’ or predictable.

$$H(P) = -\sum_{x} P(x) \log P(x)$$

OUR CASE

- High entropy ⟶ Parties talk about different topics in the same way
- Low entropy ⟶ Parties are polarized and develop more particular topics

ENTROPY


Comparison of Entropy between Political Parties and Newspapers
MF-Milano Finanza
Fratelli d'Italia
Partito Democratico
Italia Viva
Sinistra Italiana
Forza Italia
MoVimento 5 Stelle
Il Sole 24 ORE
Fanpage.it
Avvenire
Europa Verde - Verdi
Azione
Libero
Il Dubbio
Il Foglio
Il Fatto Quotidiano
RaiNews
il manifesto
Matteo Salvini
Domani
La Verità
la Repubblica
Corriere della Sera
Il Giornale
Tgcom24
La Stampa
Sky tg24
ANSA.it
Tg La7
Entity
0
1
2
Entropy
gruppo
Newspaper Headline
Political party

KL DIVERGENCE

With Kullback-Leibler divergence, it is possible to check where polarization exists.

The KL divergence measures the difference between two probability distributions.
For example between the distribution of topics covered by one newspaper and the overall distribution of topics covered by all newspapers. This helps to understand whether a newspaper focuses on certain topics disproportionately to others, indicating a possible bias.

$$D_{KL}(P||Q) = \sum_x P(x) \log\left(\frac{P(x)}{Q(x)}\right)$$

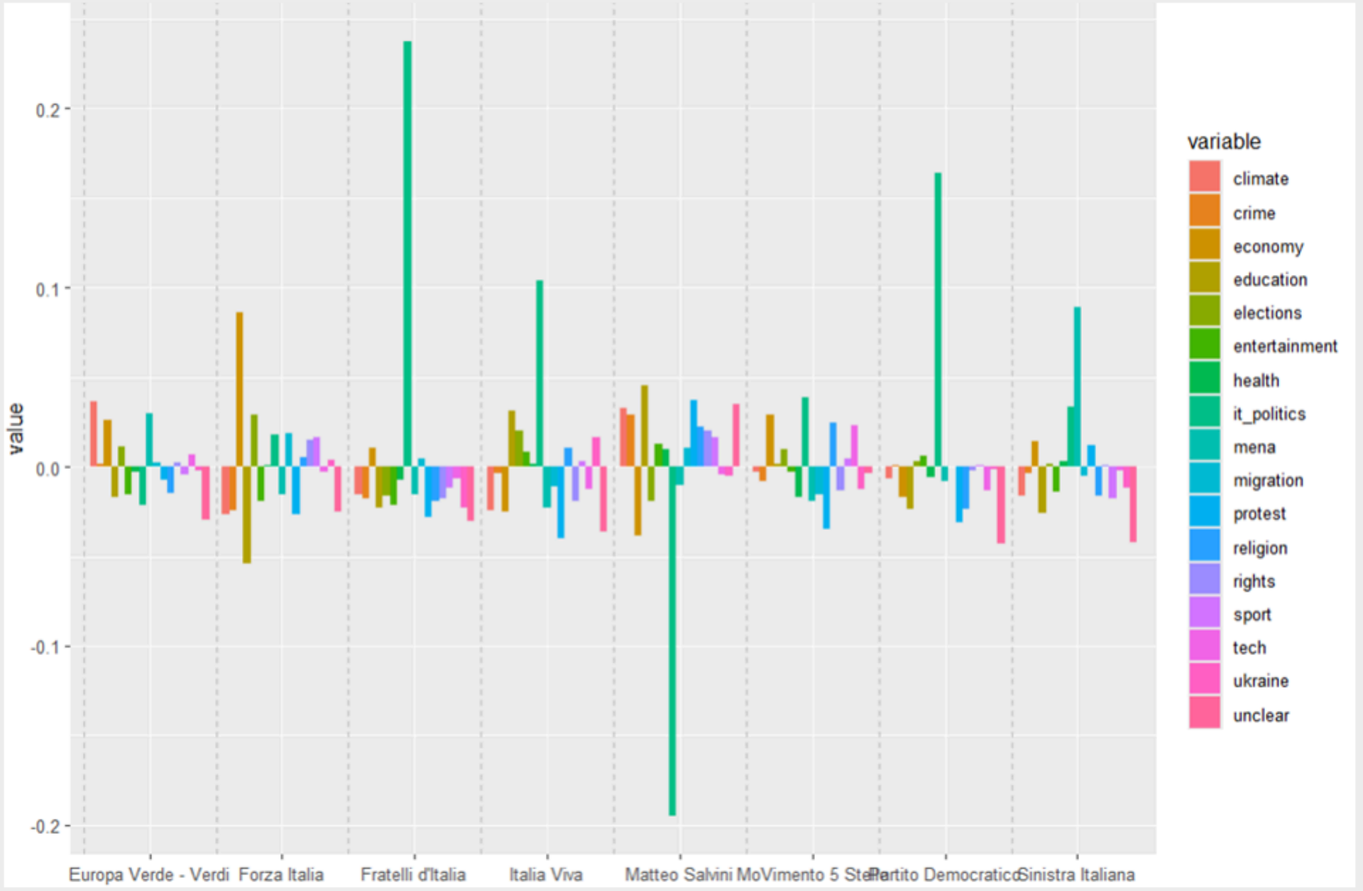
Where:
P(x) is the probability of the 'true' or 'reference' distribution P.
Q(x) is the probability of the 'specific' distribution Q.

If D = 0, the two distributions P and Q are identical.
The larger D is, the more polarised the distribution Q is from P.

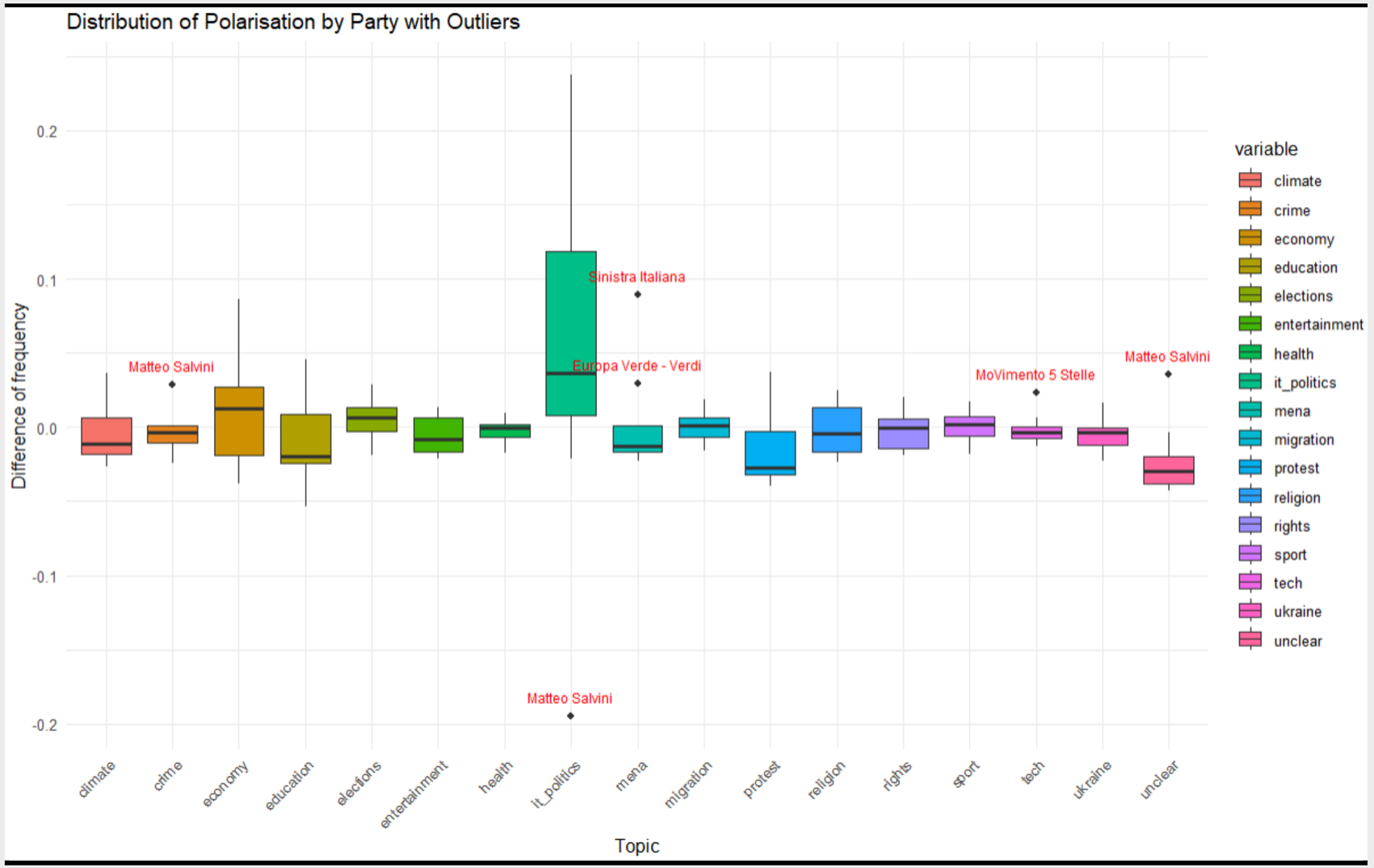
POLARIZATION IN PARTIES

We check the Z score of the parties and exclude those that have |z|<3 because they are not polarized. Then we study the distribution of the others to check which parties deviate the most from the general distribution.

Party	Z_score
Azione	-1.177908
Europa Verde - Verdi	-8.431953
Forza Italia	-5.554887
Fratelli d'Italia	-5.295193
Italia Viva	-6.991766
Matteo Salvini	-4.668333
MoVimento 5 Stelle	-8.039089
Partito Democratico	-7.246298
Sinistra Italiana	-6.447860

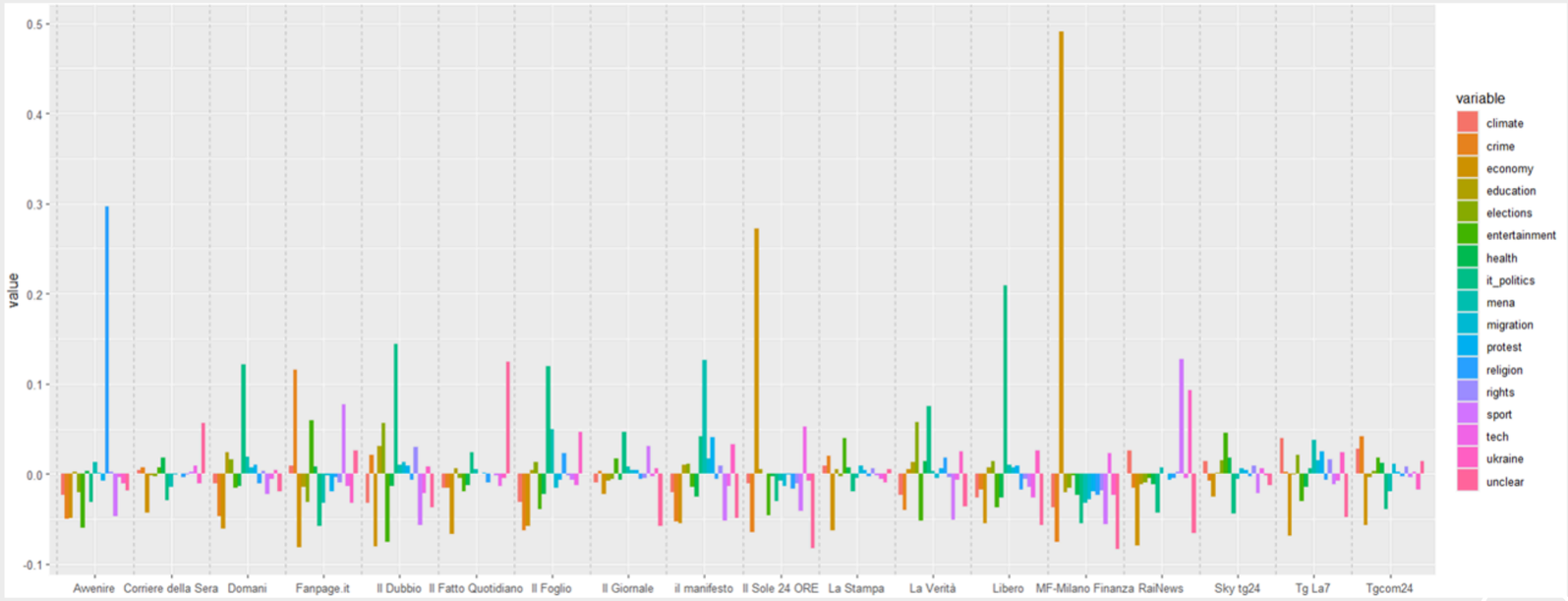


POLARIZATION IN PARTIES

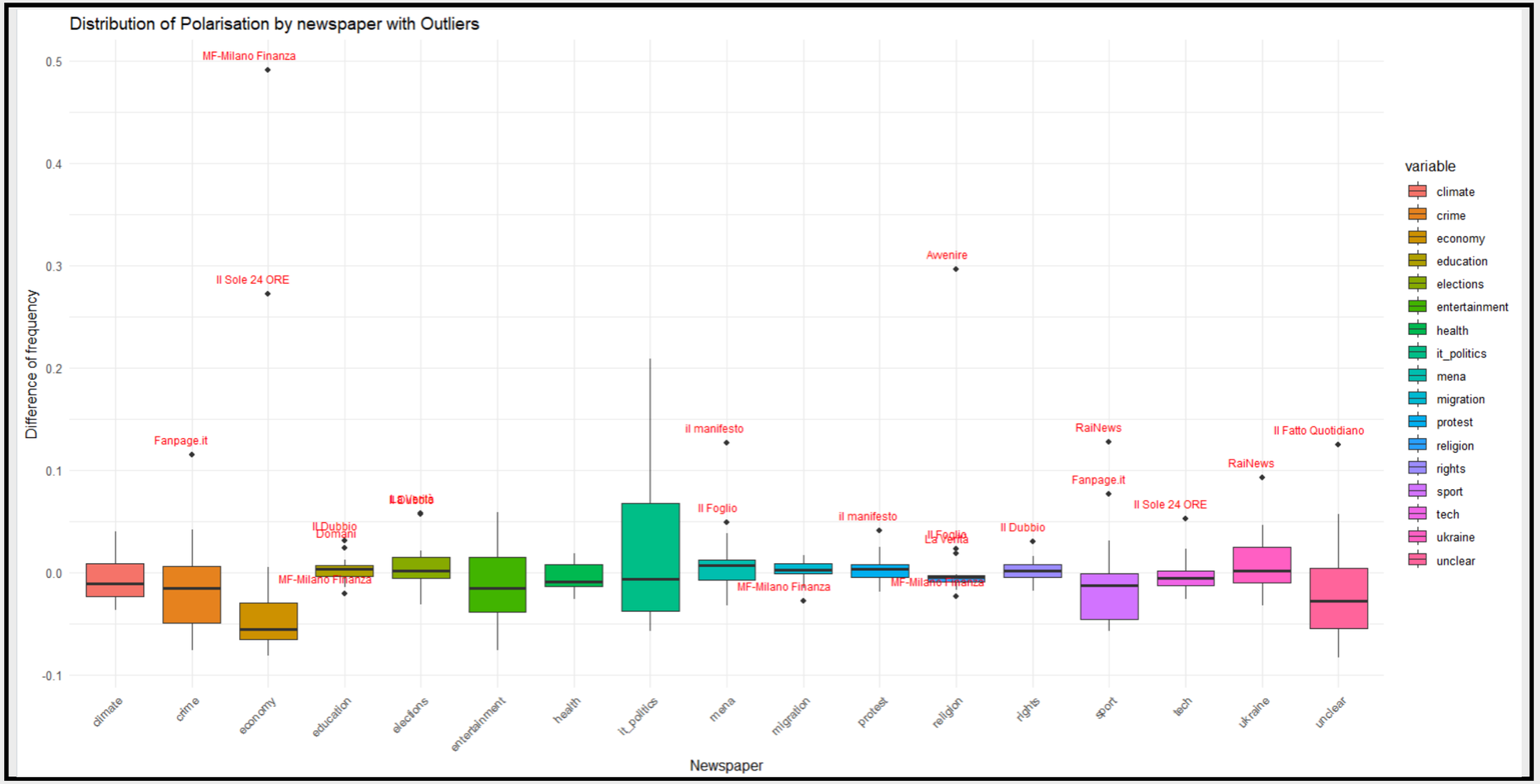


POLARIZATION IN NEWS

Newspaper	Z_score
ANSA.it	-0.274631
Avvenire	130.481977
Corriere della Sera	-12.794205
Domani	12.947274
Fanpage.it	42.449380
Il Dubbio	57.386020
Il Fatto Quotidiano	4.280485
Il Foglio	32.766023
Il Giornale	-10.575262
il manifesto	42.601840
Il Sole 24 ORE	78.478642
la Repubblica	-2.456391
La Stampa	-12.189619
La Verità	14.068762
Libero	50.135780
MF-Milano Finanza	201.906602
RaiNews	41.334376
Sky tg24	-11.494252
Tg La7	4.138948
Tgcom24	-5.988385



POLARIZATION IN NEWS



CONCLUSIONS

Polarization exists and is mainly present in newspapers.
In some cases it is justified by a declared tendency of the newspaper towards a specific topic. In other cases by an uncommunicated choice. This certainly offers many insights into whether these choices are made on purpose or not.

Crime	Economy	Education	Elections	Mena	Protest	Religion	Rights	Sport	Tech	Ukraine
Matteo Salvini	MF	Il Dubbio	La Verità	Sinistra Italiana	Il Manifesto	Avvenire	Il Dubbio	Rai News	Movimento 5 Stelle	Rai News
Fanpage.it	Il Sole 24 Ore	Domani	Il Dubbio	Verdi		Il Foglio		Fanpage.it	Il Sole 24 Ore	
				Il Manifesto		La Verità				
				Il Foglio						

SAPIENZA
UNIVERSITÀ DI ROMA

THANK YOU

Presented by Marco D'Ercole
Advisor: Walter Quattrociocchi