GreeSE Papers

Hellenic Observatory Discussion Papers
on Greece and Southeast Europe

Paper No. 210

From Ostensible to Actual Media Pluralism. An examination of content diversity in Greece's fragmented online media system.

Yani Kartalis

LSE THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

August 2025

Hellenic Observatory Centre for Research on Contemporary Greece and Cyprus

Research at LSE

Hellenic Observatory Centre
for Research on Contemporary
Greece and Cyprus
Research at LSE

From Ostensible to Actual Media Pluralism. An examination of content diversity in Greece's fragmented online media system.

Yani Kartalis

GreeSE Paper No. 210

Hellenic Observatory Papers on Greece and Southeast Europe

All views expressed in this paper are those of the authors and do not necessarily represent the views of the Hellenic Observatory or the LSE
© Yani Kartalis

LSE
THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Contents

1. Introduction 4
2. Gauging Content Pluralism in the Media 6
3. The Impact of Ownership 7
4. Data, Methods and Findings 9
4.1 Mapping the Greek Online Media 9
4.2 Diversity in Broad Topic Categories Coverage 12
4.3 Measuring Content Diversity Using Large Language Models (LLMs) 14
4.3.1 News Polarity 14
4.3.2 Pro and Anti-Government Stance 17
4.3.3 Mapping the Content of Greek News Websites 19
4.3.4 Semantic Embeddings 21
4.4 News Agency Content Reproduction 24
4.5 The Impact of Ownership 26
5. Conclusions 28
6. References 31

From Ostensible to Actual Media Pluralism. An examination of content diversity in Greece's fragmented online media system

Yani Kartalis

ABSTRACT

This study examines the Greek media landscape, characterized by a paradox of low press freedom and high numerical fragmentation. The analysis is based on an extensive dataset comprising over 3.37 million article titles from a broad spectrum of Greek online media published throughout 2024. The findings reveal that this quantitative fragmentation does not necessarily lead to qualitative content pluralism. Thematic coverage across websites shows a lack of divergence while metrics of news polarity and stance towards the government follow similar trends. Only a few websites are positioned outside the larger clusters around these metrics with the majority being pro-government and positive. News agency content reproduction further contributes to homogenization. Crucially, the study provides empirical evidence that common ownership among media outlets significantly reduces the content's diversity of viewpoints and stances. This suggests that media ownership structures exert a considerable influence on content, potentially undermining true media pluralism in Greece.

Keywords: Greek Media, Media Pluralism, Content Diversity, Media Ownership, Computational Content Analysis

Author biography: Yani Kartalis is a Postdoctoral Researcher and Data Steward at University of Copenhagen's Center for Social Data Science (SODAS). His research interests lie in the fields of representation, parliaments, political parties, EU politicization, and social data science more broadly. He has published research using a multitude of parliamentary data and data science methods to examine the ways representatives use parliamentary tools to represent their constituents in a variety of countries.

Acknowledgments: The research in this paper was possible with funding by the A.C. Laskaridis Charitable Foundation and Dr Vassili G. Apostolopoulos at the Hellenic Observatory of the London School of Economics and Political Sciences for the project titled "From Ostensible to Actual Media Pluralism. An examination of content diversity in Greece's fragmented online media system". The author would like to particularly thank PhD candidate Petros Petrousis and Professor Ioanna Kostarella at Aristotle University in Thessaloniki's School of Journalism and Mass Media Studies, as well as Dr Paris Aslanidis and the wonderful audience at the National Center for Social Research (EKKE)'s research seminars for having me present my preliminary findings and provide excellent feedback. Finally, Jonas Skjold Raaschou-Pedersen at PhD candidate at SODAS, University of Copenhagen for his help is setting up the data collection process early in the project's life cycle.

1. Introduction

The Greek media landscape currently presents an apparent paradox of media (un)freedom. While the country consistently ranks the lowest in the EU for press freedom, according to *Reporters Without Borders* (Reporters sans frontières - RSF) for three consecutive years in 2022, 2023, and 2024,[1] its media landscape appears highly fragmented, with a large number of media outlets, particularly in the digital sphere. Digitalization has further accelerated this trend, allowing independent media platforms to emerge. The diverse range of (digital) sources are used more frequently than in other countries. This is driven by the sheer number of digital outlets[2] and the widespread use of social media platforms like *Facebook, Instagram, and TikTok* for casual and incidental news consumption (Papadopoulou & Angelou, 2024).

According to the Reuters Institute's Digital News Report 2023[3], this fragmentation is not limited to digital media. The Greek television industry includes nine national broadcasters—both commercial and public—yet many struggle to attract large audiences. Meanwhile, print media, despite its decline, still features approximately 20 national daily newspapers and seven sports newspapers, though most have limited readership and lack financial sustainability. Additionally, the persistent absence of public data on online advertising market shares and audience shares of top digital players hinders a precise evaluation of market concentration and its effects on media pluralism (Newman, et. Al., 2024; Papadopoulou & Angelou, 2024).

At first glance, the abundance of Greek media outlets—especially digital ones—suggests a pluralistic and diverse information ecosystem. But the apparent numerical fragmentation of the media system in Greece does not guarantee a translation into actual pluralism in content and viewpoints. Media pluralism is not simply a matter of counting outlets; it also concerns the diversity of perspectives presented across and within media enterprises. Pluralism requires ensuring there is a diversity of viewpoints that prevent any one (or few) media owner(s) or voice(s) having too much influence over public opinion and the political agenda.

---

[1] https://rsf.org/en/index

[2] The now-outdated Greek General Secretariat for Media and Communication's online media registry, "e-media" alone used to catalogue nearly 1,500 voluntarily registered digital media platforms. As of 2022, this platform has given its place to the current registry. The website is still live at the time of writing these words at https://emedia.media.gov.gr/certified-enterprises

[3] Reuters Institute for the Study of Journalism - University of Oxford - Digital News Report 2023 - Greece - https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2023/greece

This is equally important to nominally having multiple outlets available.

Scholars studying the Greek media ecosystem have already reported a "low level of independence and pluralism in journalism, regenerating a culture receptive to the model of media instrumentalization and governed by self-censorship when reporting news" (Papathanassopoulos, et. al., 2021). Similarly, the European University Institute's Media Pluralism Monitor (MPM2023) identifies Greece's market pluralism as being at high risk, largely due to the influence of media ownership and commercial interests over editorial independence (Papadopoulou & Angelou, 2024). Papadopoulou and Angelou tell us how the existing literature has documented that the political stance of a medium's owners in Greece is still the defining factor behind its news framing.

At the heart of the issue is significant media capture, with private media dominantly owned by a handful of wealthy families and entrepreneurs with far-reaching interests in sectors like shipping, finance, and energy. This nexus of business and political elites shapes the content and orientation of much of the country's media, with major outlets often aligned with specific political agendas.[4]

Furthermore, low trust in media which follows a broader pattern of institutional distrust in Greece[5], raises further concerns about the role of the press in democratic discourse. Media play a crucial role in sustaining the quality of democracy, and these concerns about media freedom in Greece have sparked debates about whether such freedom is under threat and what implications this may have for democratic discourse. Despite the intensity of these discussions, we lack the empirical evidence to support or refute claims about the relationship between media fragmentation and actual content pluralism. Does the sheer number of Greek (digital) media outlets result in a genuine pluralism of opinions and perspectives? Or is the fragmentation merely quantitative, with certain viewpoints being disproportionately represented while others struggle for visibility? Additionally, and perhaps more importantly, how does media ownership influence content diversity?

This study seeks to contribute empirically to this debate by systematically mapping the content of Greek digital news media. By analyzing the range of perspectives, sentiment, stances, and topics covered across digital media outlets it aims to assess whether the observed

[4] The International Press Institute (IPI)'s Media Capture Monitoring Report: Greece (2024). Available at https://ipi.media/publications/media-capture-monitoring-report-greece/

[5] Reuters Institute for the Study of Journalism - University of Oxford - Digital News Report 2023 - Greece - https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2023/greece

quantitative fragmentation corresponds to a qualitative one. Furthermore, it looks at how ownership affects this qualitative fragmentation.

The paper is structured as follows: first, it outlines the theoretical framework and existing literature on media pluralism and ownership influence. Second, it presents the methodology employed to assess content diversity across Greek news media. Third, the findings will be discussed in relation to quantitative and qualitative measures of pluralism. Finally, the conclusion reflects on the implications of these findings for Greek media freedom and democratic discourse.

2. Gauging Content Pluralism in the Media

The assessment of media content pluralism is crucial for evaluating the diversity of information available to the public. Studies have already approached the problem of empirically assessing content pluralism in the media. A range of methodological approaches to measure pluralism have been employed, reflecting different theoretical and practical considerations. These approaches span from qualitative content analysis to advanced computational techniques, each with its own advantages and limitations.

Some studies approach the matter with a normative lens, implementing multi-perspective frameworks (Udris, 2023; Magin et al., 2023). They incorporate journalistic standards, considering additional factors such as neutrality, discursivity, and civility, which contribute to a more holistic understanding of media quality. While these frameworks can provide valuable qualitative insights, their operationalization often requires subjective coding, which may limit replicability across contexts. Other approaches focus on source tracking methodologies to assess media diversity through retroactive source auditing (Grannan et al., 2025).

One of the most prominent frameworks for assessing media pluralism in Europe is the Media Pluralism Monitor (MPM), developed by the European University Institute. The MPM provides a holistic assessment of risks to media pluralism through a set of twenty indicators, covering legal, economic, political, and socio-cultural dimensions. The 2023 edition (Bleyer-Simon et al., 2024) incorporates measures of ownership concentration and network structures, allowing for a nuanced understanding of how content is distributed across digital platforms.

A growing body of research employs multi-platform content analysis to examine patterns of

content distribution across digital and traditional media. They investigate how news is bundled and repurposed across platforms, revealing potential risks of content homogenization and agenda-setting influence by dominant actors (Champion, 2015).
More recently, the integration of computational methods, such as network analysis and topic modelling, has expanded the capacity to analyze large-scale datasets and uncover hidden structures of media influence (among others: Blair, et. Al., 2020, Pinto, et. Al., 2019). Additionally, sentiment analysis has emerged as a valuable tool for evaluating the ideological and emotional tone of news content. Researchers employ lexicon-based approaches, supervised machine learning models, and transformer-based model architectures like BERT and GPT to capture sentiment shifts in political and economic news coverage (e.g. Bashiri & Naderi, 2024).

3. The Impact of Ownership

Existing research on media ownership consistently points to the influence of ownership concentration on news content. Several scholars argue that ownership structures shape editorial decision-making processes, which in turn affect how widely and deeply different perspectives are represented (e.g. Baker, 2006; Champion, 2015). Broadly, the literature underscores both potential benefits and threats to pluralism stemming from different ownership configurations. The normative assumption is that increased market concentration and monopolization may have a negative impact on journalistic autonomy, news diversity, independence, and democracy (Sjøvaag and Ohlsson, 2019).
According to Noam (2016), concentrated media ownership is one of those fundamental issues of power and wealth distribution that every generation must resolve. However, Trappel & Meier (2022) contend that three decades of regulatory efforts at the EU level have failed to dismantle oligopolistic media structures as member states decide to often shield legacy media from digital disruption without curbing their potentially undemocratic practices.
A key theme emerging in the literature is the link between increased ownership concentration and reduced content diversity. Multiple works show us how chain ownership formation (a single company or individual owning multiple media outlets) increases homogenization (Baum & Zhukov, 2019; Hendrickx & Ranaivoson, 2021; Vogler et al., 2020). Private owners are primarily economic actors, pursuing economic profit, which in turn

marginalises less popular and consequently expensive content and prioritizing profitable content (Korbiel & Sarikakis, 2017). Hanretty (2014), drawing on insights from 32 European media markets, similarly finds that owners owning multiple outlets at the national level only are more likely to exercise owner influence.

Despite this overarching concern, recent research suggests that ownership concentration does not uniformly diminish pluralism. In their study, Hendrickx and Kammer (2024), show that corporate strategies could engage in both synergy and diversification practices in the context of large media-chain groups. Furthermore, in other contexts, chain-owned newspapers were found to exhibit greater political viewpoint diversity than single-owned publications, indicating that economies of scale can, under certain conditions, foster pluralism (Garz et al., 2024) while also potentially increasing certain aspects of content quality (Garz & Ots, 2025).

Another recurring theme is the importance of editorial independence in maintaining diversity within individual news outlets. Even when ownership is concentrated, strong editorial autonomy can allow for a broader spectrum of opinions, mitigated corporate pressures, and healthier representation of local issues (Hendrickx & Kammer, 2024). Additionally, robust regulatory frameworks can hold concentrated ownership entities accountable for providing balanced news coverage (Hendrickx & Remoortere, 2024). At the same time, Champion (2015) contends that emerging digital platforms may offer new avenues for diverse content creation, potentially counteracting some of the negative effects associated with ownership concentration. Audience fragmentation, coupled with lower barriers to entry for online publishing, creates opportunities for alternative voices to flourish.

Overall, the scholarship highlights the complex role that ownership structures play in shaping content pluralism. While ownership concentration poses clear risks for diversity and editorial independence, digital technologies can provide avenues for more perspectives. Consequently, effective regulatory frameworks and continuous monitoring of both traditional and digital outlets are essential to ensure that media pluralism is preserved despite ongoing changes in ownership patterns.

Against this backdrop, Greece presents a very informative case. Beyond the severe concentration of media ownership, the country's dynamic digital media field could be offering more pluralism than expected. This study presents a comprehensive empirical

examination of the content in the Greek online media sphere, employing a range of quantitative methods to assess media pluralism and diversity. Building upon established approaches in news content analysis—such as topic distribution, sentiment analysis, and stance detection toward the government— the study also integrates advanced natural language processing (NLP) methodologies and novel operationalizations.

By doing so, it seeks to expand upon existing literature and provide a more granular understanding of the structure of Greek online news content. This is an empirical contribution to a widely held but underexamined expectation regarding the diversity of perspectives in the Greek media ecosystem. One that assumes media fragmentation (quantity of outlets) leads to content pluralism (quantity of opinions).

4. Data, Methods and Findings

The following sections outline the methodology and key findings of an extensive empirical analysis of content within the Greek online media landscape. Building on established approaches to news content analysis—such as topic distribution, sentiment analysis, and stance towards the government—this study incorporates advanced natural language processing (NLP) techniques to quantitatively gauge content plurality.

Adopting a quantitative lens, the research pursues three primary objectives: (a) to empirically investigate the relationship between media fragmentation and content pluralism; (b) to assess the diversity of topics, sentiments, and viewpoints across Greek news outlets; and (c) to explore how shared ownership structures may influence content diversity. The analysis is based on an exhaustive and very broad cross-sectional data collection effort which leads to a large-scale dataset comprising articles from the full spectrum of Greek online media published throughout 2024.

4.1 Mapping the Greek Online Media

Greece's chaotic digital media ecosystem makes any effort to systematically map and research the field and the ownership patterns very difficult. The main starting point in terms of data collection is the Greek Electronic Press Register (*Μητρώο Ηλεκτρονικού Τύπου - ΜΗΤ*)[6] of the General Secretariat of Communication and Information. As of 2022[7], electronic

[6] https://mt.media.gov.gr/submissions/MHT/public/

[7] The current Register replaced the previous iteration that was introduced in 2015.

press enterprises[8], that wish to be certified, need to apply in September of each year. By joining the Register, each website acquires a unique registration number and a mark certifying its registration. Over 515 news websites are currently[9] registered. Some websites are still not registered and are added manually to the list of websites either because they are deemed important or because they have some known ownership connection with some other website in the dataset.[10] This brings the total number of websites in the dataset to 531. Out of those, 275 (52.4%) are local websites, mostly containing news about some specific region or city. Figure 1 shows the geographical distribution of these local websites across regions in Greece.

Not all websites on MHT are news websites. Some carry content mostly on economic matters while others are monothematic and only publish articles on categories such as sports, leisure activities, health etc. Figure 2, shows the distribution of websites in the dataset across these broader content categories. There are 90 websites (17.2%) about general national news and 30 (5.5%) about Economy and Business news. Beyond these two, other major categories include Lifestyle, and Sports which together comprise a total of 41 websites (7.7%). For the purposes of the analysis presented in the following sections, the study focuses on the 120 *General News,* and *Economy and Business* websites (a combined 22.7% of websites).

[8] As defined in paragraph II of Article III of Law 5005/2022.

[9] As of 15 November 2024.

[10] Those include: thepressproject.gr, neolaia.gr, pentapostagma.gr, hellasjournal.com, euronews.com, iellada.gr, ertnews.gr, podcastmedia.gr, flight.com.gr, e-amyna.com, sdna.gr, sport-fm.gr, monobala.gr, skaikairos.gr, agrocapital.gr, agro24.gr, akispetretzikis.com

Figure 1. Geographical distribution of registered local websites per region – NUTS2

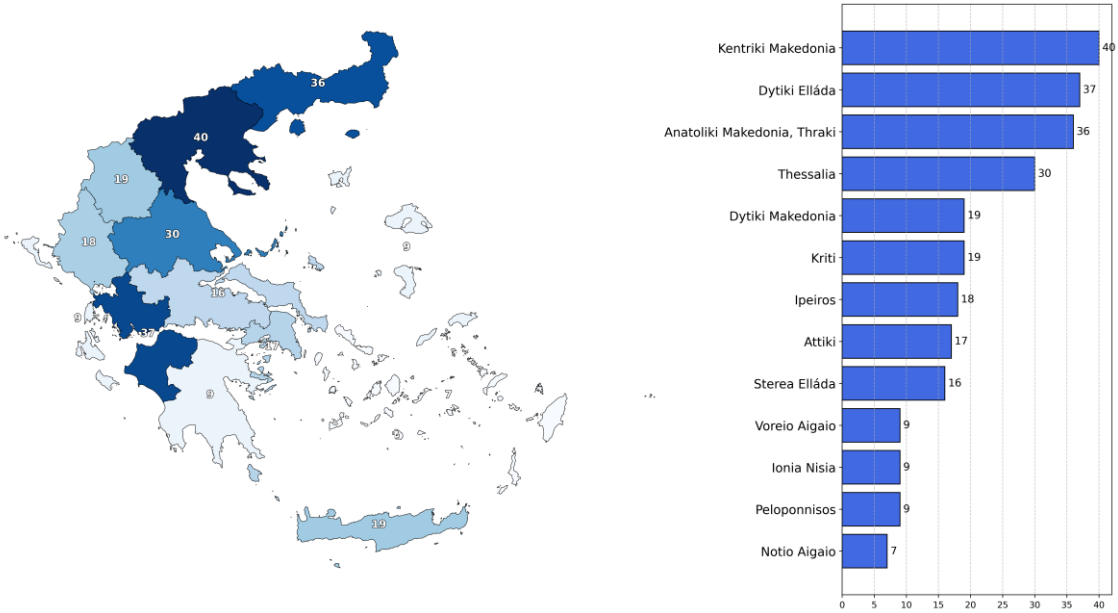
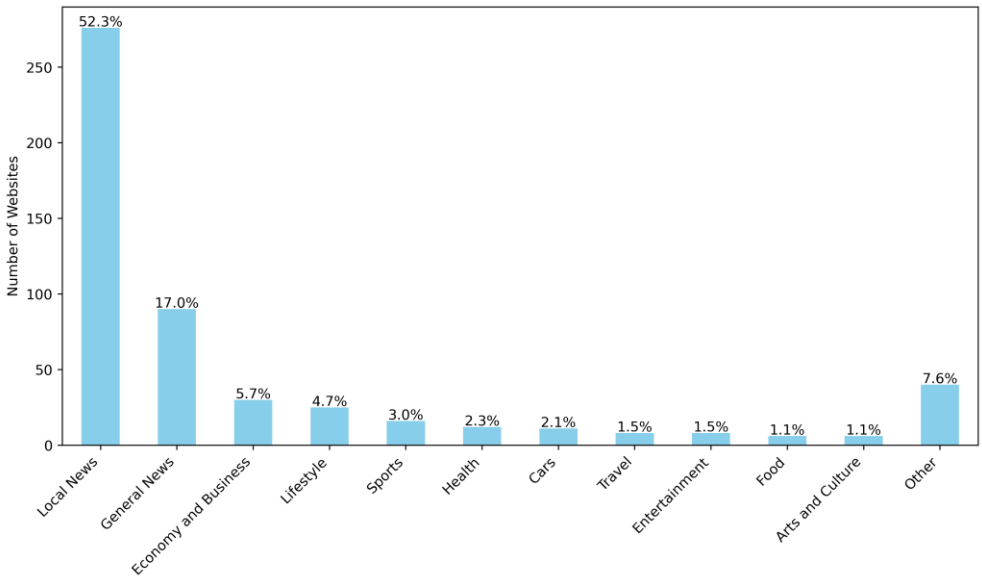


Figure 2. Distribution of registered websites per overall news category

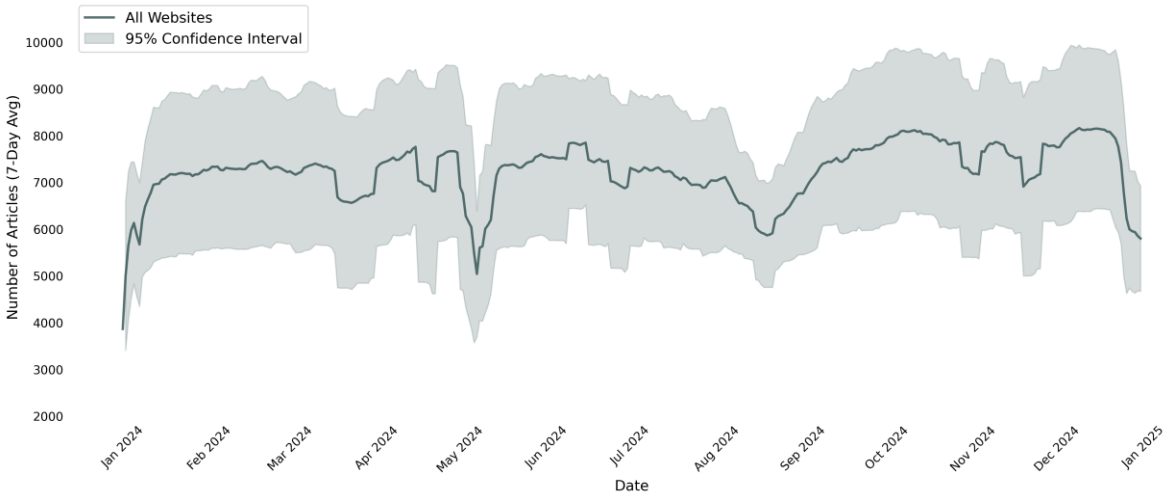


In order to collect all the articles a website publishes, the study turns to what is publicly available in each page. Out of the 120, due to various technical limitations on the side of the websites[11], 81 had all their articles readily and freely available for collection at the time of writing this piece. For those, **all** article titles published in 2024 (1/1/2024 – 31/12/2024) are collected. A complete list of websites for which data is collected is available in Table A2 in

[11] Some websites do not offer publicly available data per date, beyond what has been published on the actual day. Other websites do not go far back enough or crash while loading previous content.

the appendix.[12] A grand total of 3,379,723 article titles are included in the study. Figure 3 shows the temporal evolution of article production from all the websites in the dataset.

Figure 3. 7-Day Rolling Average of Articles Per Day (01.01.2024-31.12.2024)



This large dataset is annotated automatically on a series of metrics that together attempt to capture content pluralism in the Greek online media sphere. The following sections lay out this extensive annotation process that is undertaken to analyse various dimensions of media content. This effort includes sentiment analysis, categorization according to the International Press Telecommunications Council (IPTC) standards, assessment of pro- or anti-government stances, and the generation of content embeddings using latest advances in NLP.

4.2 Diversity in Broad Topic Categories Coverage

First, article titles are classified based on International Press Telecommunications Council (IPTC) standards, which provide a comprehensive framework for categorizing news content. These categories are: *1) arts, culture, entertainment and media, 2) conflict, war and peace, 3) crime, law and justice, 4) disaster, accident and emergency incident, 5) economy, business and finance, 6) education, 7) environment, 8) health, 9) human interest, 10) labour, 11) lifestyle and leisure, 12) politics, 13) religion, 14) science and technology, 15) society, 16) sport, 17) weather*. Additionally, *18) local news, 19) international*, and *20) uncategorized*

[12] Notable exceptions are *protothema.gr* and *real.gr,* as well as *parapolitika.gr* for which data only goes back to November 2024. Numerous communications with management and technical departments led to no avail. The newspapers either never responded or refused to grant access to their article collection. Other exceptions include: *cnn.gr, ant1news.gr, huffingtonpost.gr, elora.gr, tvopen.gr*. In stark contrast, data for AMNA (ΑΠΕ-ΜΠΕ ΑΕ) the Greek news Agency, was sent fully annotated for the purposes of the study after signing a research data access agreement.

were added to these categories to account for some of the article titles that did not belong to any of these.

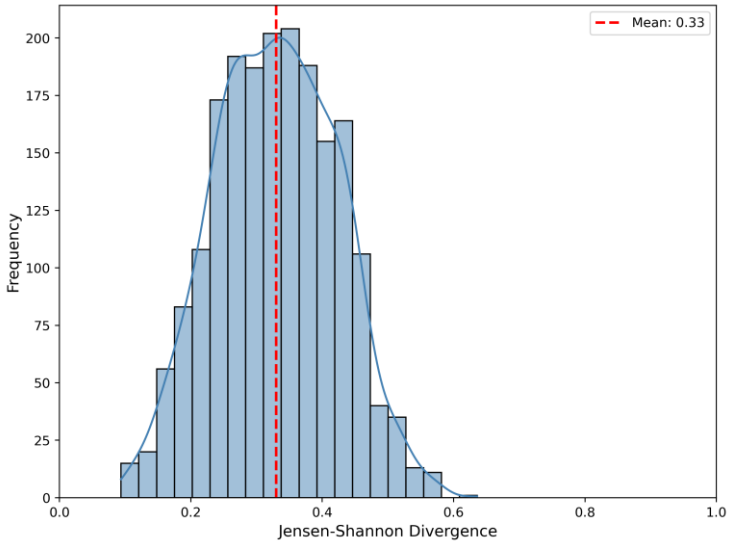
This classification is done initially by taking advantage of each website's available information for each article. A vast majority of articles contained some kind of topic category as a tag, or in their *url*. These were assigned to each one of the IPTC categories manually. The remainder (about 20% of the article titles) were annotated by means of a zero-shot annotation using OpenAI's GPT-4o model.[13] Table A1 in the appendix provides detailed descriptions for each IPCT category. This categorization facilitates a structured analysis of topics and themes prevalent in the Greek online media landscape in 2024.

To estimate the diversity in coverage between websites in these topic categories a divergence score is calculated. The aim is to compare the thematic content distribution by looking at how similar or dissimilar their coverage is across the predefined IPTC news categories. The measure calculates the relative frequency of articles in each category for each website, constructing a set of probability distributions. Pairwise Jensen-Shannon Divergences are computed between these distributions for all website pairs, providing a measure of content dissimilarity between websites. Jensen-Shannon Divergence (JSD) is a common metric in the literature (e.g. Blair, et. Al., 2020)

Figure 4 presents the distribution of these pairwise JSD scores, with a vertical line marking the average divergence (system-level), offering a concise overview of how thematically similar or diverse the sampled websites are in their content focus. At the system level, the mean JSD score of *0.33* means that the Greek online media sphere is non-divergent. This means that the average distance between the distributions of topic coverage for all pairs of websites is small and the websites are thematically similar overall, following the same patterns and fluctuations. Normatively speaking, this is not necessarily something negative as it could simply mean that the websites pay attention to the same broad categories as news take place and require coverage.

[13] This annotation step was validated by using the data furnished by AMNA. This rich data contained IPTC annotations that were used in contrast with the model's annotation. This process yielded an accuracy score of 0.94.

Figure 4. Distribution of Jensen-Shannon Divergence for all Website Pairs



4.3 Measuring Content Diversity Using Large Language Models (LLMs)

Next, a series of LLM-annotated content measures are implemented. For this annotation, the study only focuses on a subset of the 3.37 million article titles that contains those titles that were previously annotated as a) *Political* or b) *Economy, Business and Finance,* as per the IPTC taxonomy.[14] This subset comprises of 848431 (25%) article titles.

4.3.1 News Polarity

These article titles are annotated for their general polarity. Each title is zero-shot annotated by an LLM on whether it conveys a positive or negative sentiment on a scale of -1 (Very Negative) to 1 (Very Positive).[15] Figure 5 shows histogram distributions of all title sentiments for each website, ranked by their average polarity score.

The mean, system-wide polarity score is 0.04. This indicates that most articles in the websites included in the dataset (at least in these two topic categories) are predominantly

[14] This decision is primarily made due to constraints in resources and the inherent high cost of utilizing proprietary LLMs but also due to the assumption that most notable differences in political/social content diversity are expected to be present in these two news categories and not some other (for example sports).

[15] More specifically, the GPT-4o-mini model, using the following prompt: "*You are an expert sentiment annotator of article titles from Greek news websites. Analyze the sentiment of the following title. Indicate the sentiment it conveys with regards to Greece by assigning a continuous sentiment score between -1 and 1, where -1 is the most negative, 0 is neutral, and 1 is the most positive, and provide your confidence score as a decimal between 0 and 1. Always use the following structured response format: sentiment score, confidence score. No explanations*.". Additionally, the "temperature" parameter is set to 0.01 for all model calls and the seed kept stable. A validation exercise with 300 manually annotated random sample of article titles yielded an accuracy score of 82%.

positive. Even if not by a large margin, this finding comes as a surprise as the assumption is that most news outlets and media systems nowadays choose to report negative news at a higher rate. This is an empirical finding one can find in a multitude of studies employing polarity measures (e.g. Costola, et. al., 2023; Rozado, et. al., 2022; Koch, et. al., 2024). More specifically, the same system-wide polarity score reduces to -0.07 if one only looks at articles that fall under the "*Politics*" IPCT news category. This finding aligns with the existing literature. In contrast, the same score increases to 0.17 for "*Economy, Business, and Finance*". Economic/financial news websites, are more likely to contain positive articles. This is also evident in Figure 5, where the top-ranking websites in terms of positive polarity are economic/financial ones.

Figure 5. Distribution of news polarity across all websites - ranked by average (low to high)

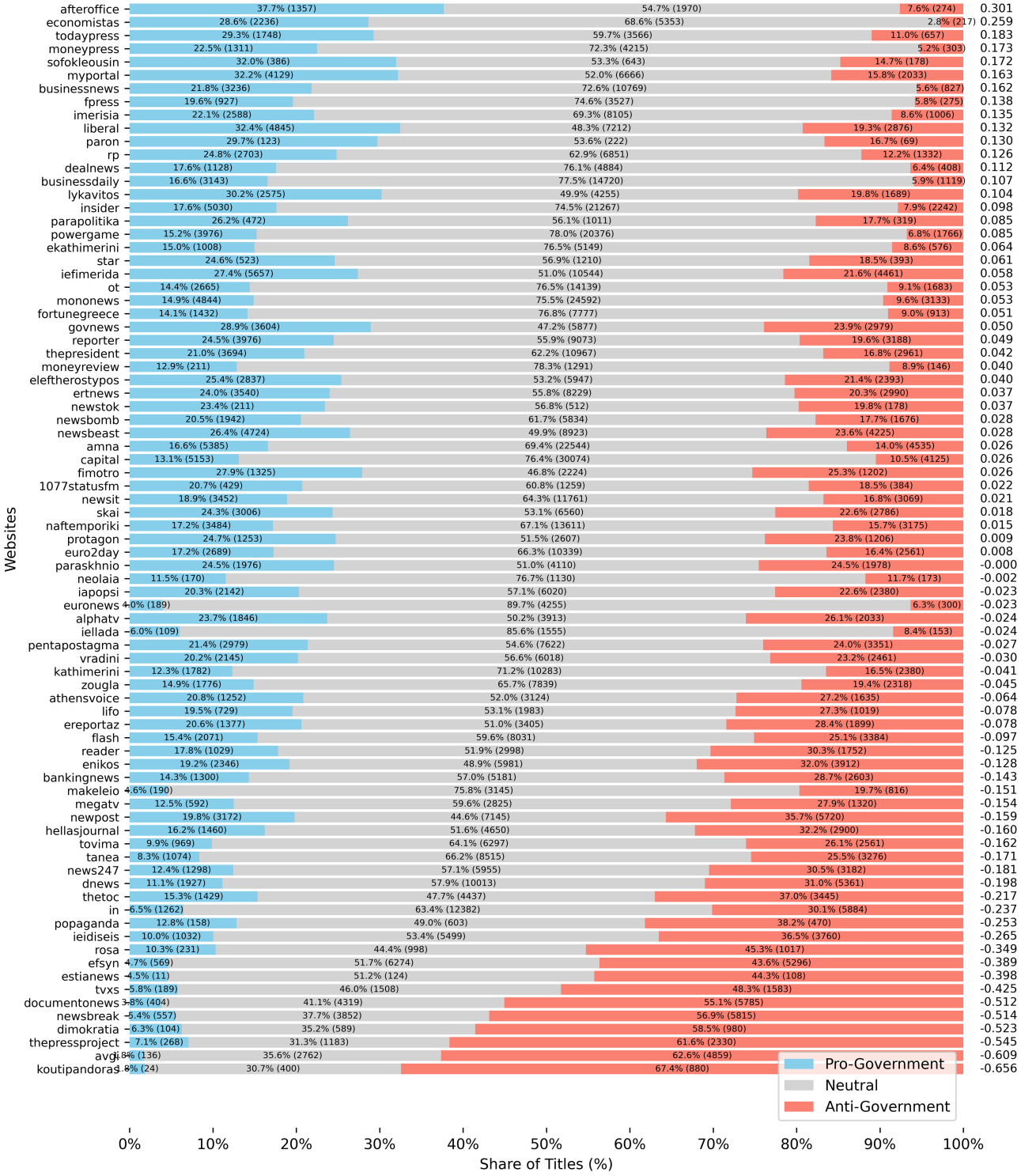
4.3.2 Pro and Anti-Government Stance

The articles are further annotated for their stance towards the government. Each article title is zero-shot annotated by an LLM on whether it conveys a positive or negative evaluation towards the government.[16] Figure 6 shows how the shares of pro and anti-government article titles are distributed for each website, ranked by the overall weighted stance towards the government.

The mean, system-wide and weighted general stance score is -0.05 (further decreased to -0.07 if the neutral articles are removed). This indicates that a slight majority of Greek online news articles (at least in these two topic categories) are anti-government. Importantly, *AMNA* and *ertnews*, the only two public outlets in the dataset, stand clearly on the pro-government camp although not too far from the middle. Not an unexpected finding given the documented operation of both outlets under the "direct control of the executive" (Papadopoulou & Angelou, 2024). However, if one looks at the number of websites that are pro-government (general weighted stance towards the government equal or greater than 0.001) more than half of the websites (43) could be considered pro-government.

[16] More specifically, the GPT-4o-mini model, using the following prompt: "*You are an expert annotator specializing in identifying pro-government and anti-government stances in article titles from Greek news websites. You are working specifically during the term of the Mitsotakis government (2019–2024). For each article title shown to you, analyze its stance and respond with one of the following: 'Pro-government' if the title supports, favors, or promotes the actions, policies, or image of the Mitsotakis government. 'Anti-government' if the title criticizes, opposes, or undermines the actions, policies, or image of the Mitsotakis government. 'Unsure' if the title is neutral, ambiguous, or lacks sufficient context to determine a stance. No explanations.*". Additionally, the "temperature" parameter is set to 0.01 for all model calls and the seed kept stable. A validation exercise with 300 manually annotated random sample of article titles yielded an accuracy score of 85%.

Figure 6. Distribution of shares of stances towards the government - ranked by average (low to high)



4.3.3 Mapping the Content of Greek News Websites

Furthermore, Figure 7, plots the two metrics for all websites on a two-dimensional plane where the y-axis shows the website's average polarity score, and the x-axis its weighted average stance towards the government. This is done in an effort to visually map the Greek online media sphere.

Figure 7. Mapping Greek Websites - Mean Polarity vs General Stance Towards the Government

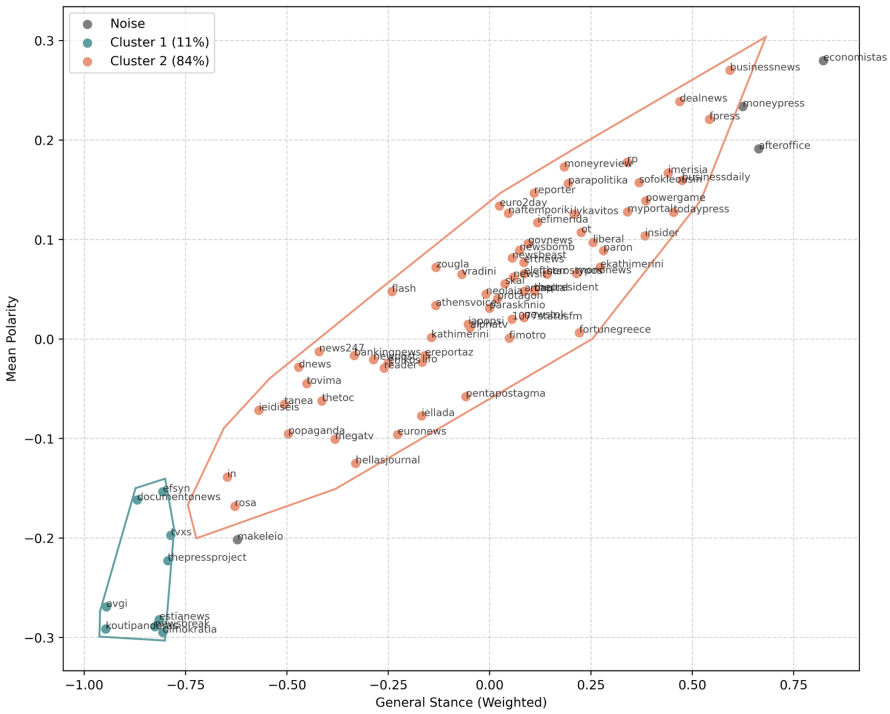
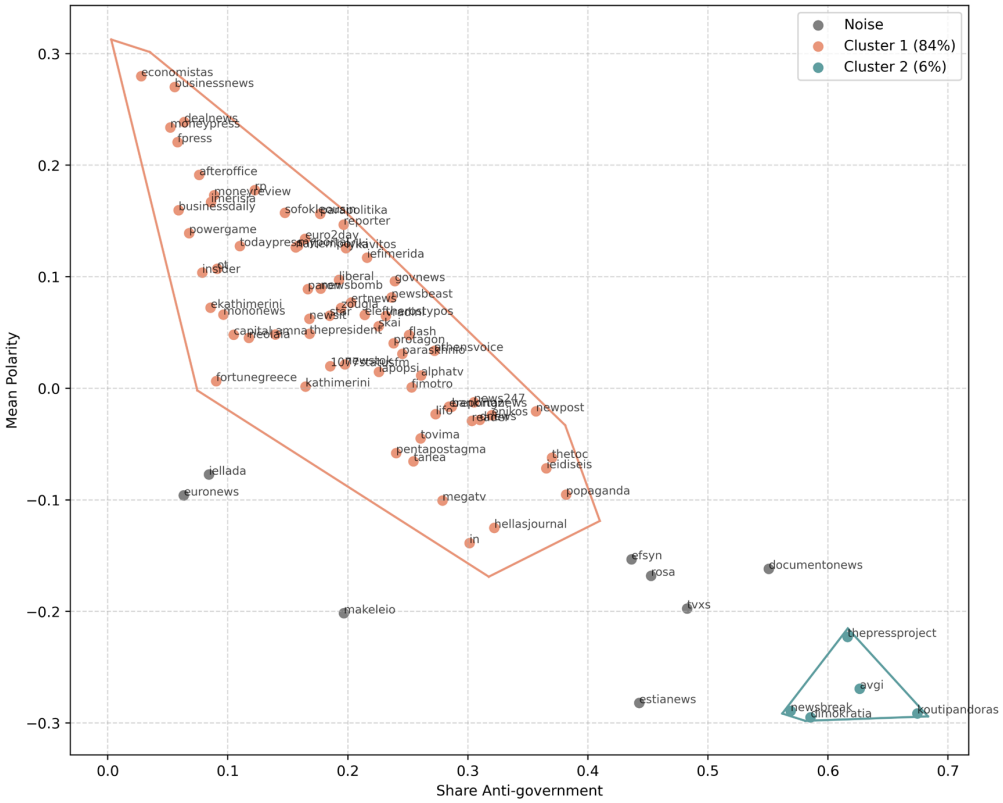


Figure 8. Mapping Greek Websites - Mean Polarity vs Share of Anti-Government Articles



Placement appears valid at face value with websites being plotted at the expected position and no stark and counterintuitive outcomes. To identify and visualize clusters among these news websites based on these metrics, an unsupervised clustering approach is implemented. Using HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise), which allows for the classification of outliers as noise, each website is represented as a labelled point that forms part of a cluster. This approach provides an interpretable spatial representation of content pluralism and alignment in the Greek digital media landscape.

An apparent clustering of websites into two main clusters is evident from this figure. One somewhat negative but pro-government and another one way more negative and anti-government. The first cluster is arguably much larger, containing 84% of websites, while the second (mostly containing left-wing, independent media) consists of only 11% of websites in the dataset.

This clustering becomes even more apparent when mean polarity is plotted together with the share of anti-government articles for each website in Figure 8. There is only a handful of websites in the sample with content severely opposing the government (share of anti-government higher than 0.4) while the vast majority of websites laying somewhere in the

region of 0 and 0.4. Apparently, not many websites report negatively on the government but those that do, do so more fiercely. This is what probably causes the overall system to look more critical of the government when looking at the overall system-wide mean values.

4.3.4 Semantic Embeddings

To further enrich the content diversity analysis, the full dataset (all 3.37 million article titles) is annotated with dense semantic embeddings. The embeddings are generated using the open-source, *paraphrase-multilingual-MiniLM-L12-v2* language model which supports multilingual text, including Greek. Sentence-level embeddings with 384 dimensions are computed for each article title, capturing their semantic meaning in a high-dimensional vector space. The assumption is that these embeddings provide a numerical representation suitable for the identification of thematic patterns across the corpus. This approach enables a more fine-grained, content-based exploration of media pluralism, moving beyond traditional keyword- and topic-based methods.

These embeddings are first averaged per website to provide one single embedding value for all their articles per IPCT category. The idea is that these mean embeddings carry the average values of all the textual (content) characteristics of all their article titles. Then each website's mean embedding for each IPCT category is compared with all other websites' equivalent scores with a simple cosine distance metric. The score is then turned into a similarity one (1-cosine distance). This ultimately produces pairwise matrixes of similarity scores across all website pairs per IPCT category. Figure 9 shows the average similarity score of all website pairs per IPCT category. Politics is by far the most similar topic at the system level meaning that websites talk about political matters in a similar way on average.

The 384-dimension mean embeddings for each website are further processed to allow for more direct comparisons between websites and their underlying approach or framing in terms of content in their news. Instead of using the mean value, a dimensionality reduction approach is implemented using uniform manifold approximation and projection (UMAP). The mean embeddings are standardized and reduced to a one-dimensional reduced representation. The resulting *Reduced Embedding (1D)* serves as a compact representation of each website's textual characteristics. This reduced representation is then plotted against the mean polarity **(**Figure 10) and stance towards the government (Figure 11) of all the articles from each website, providing a visualization of how sentiment and stance towards

the government align with the underlying content structure of the text data.

Figure 9. Average Cosine Similarity of all Pairwise Comparisons of Websites by IPCT Category

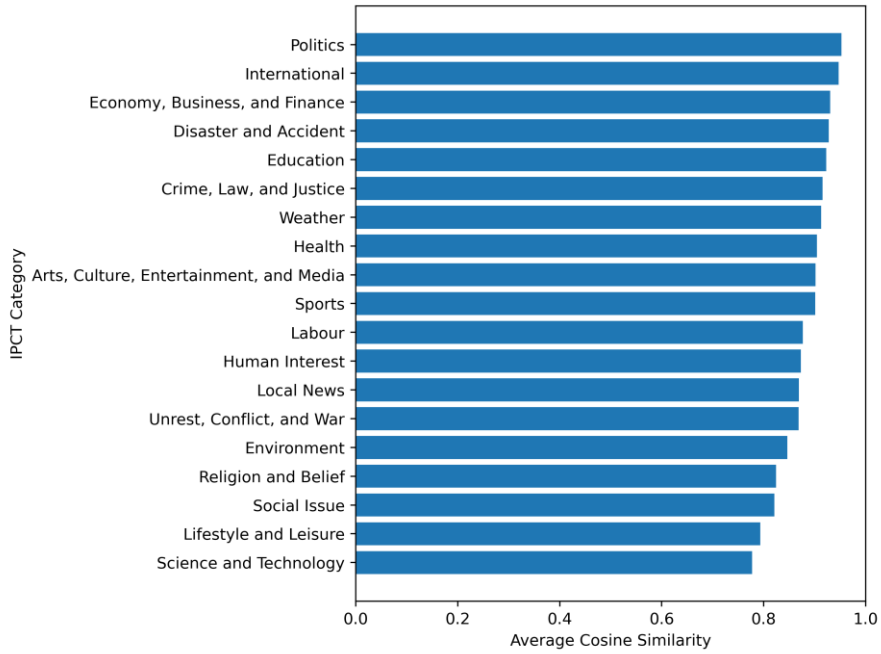


Figure 10 shows a clustering of websites into three main clusters.[17] On the one hand, in the bottom right corner of the plot, lie the economic websites which are predominantly pro-government. Then a large cluster is right on top of those which contains websites that are close to zero in terms of stance towards the government or slightly leaning towards the anti-government side (max -0.2). Finally, the third cluster which contains websites that are very anti-government (below -0,2) but speak about issues in a similar way (*Reduced Embedding – 1D)*. Similarly, Figure 11 provides a comparable snapshot. In the bottom right corner, there is a cluster of economic/financial websites that speak about matters in a very similar way (*Reduced Embedding – 1D)* and are also mostly positive in terms of polarity. Then a second very large cluster of average polarity websites (ranging between -0.05 and 0.2 with some outliers) that speaks about issues in a similar but broad way in terms of embeddings. Finally, a third cluster of very negative websites that are also speaking about the same issues in a very similar way. These two figures tell a similar story of majority of websites moving between similar lines of content with those deviating being a small minority.

[17] Automated clustering did not yield robust results like in the previous sections and these clusters are described as observed in the plot.

Figure 10. General Stance Towards the Government vs Reduced Embedding (1D)

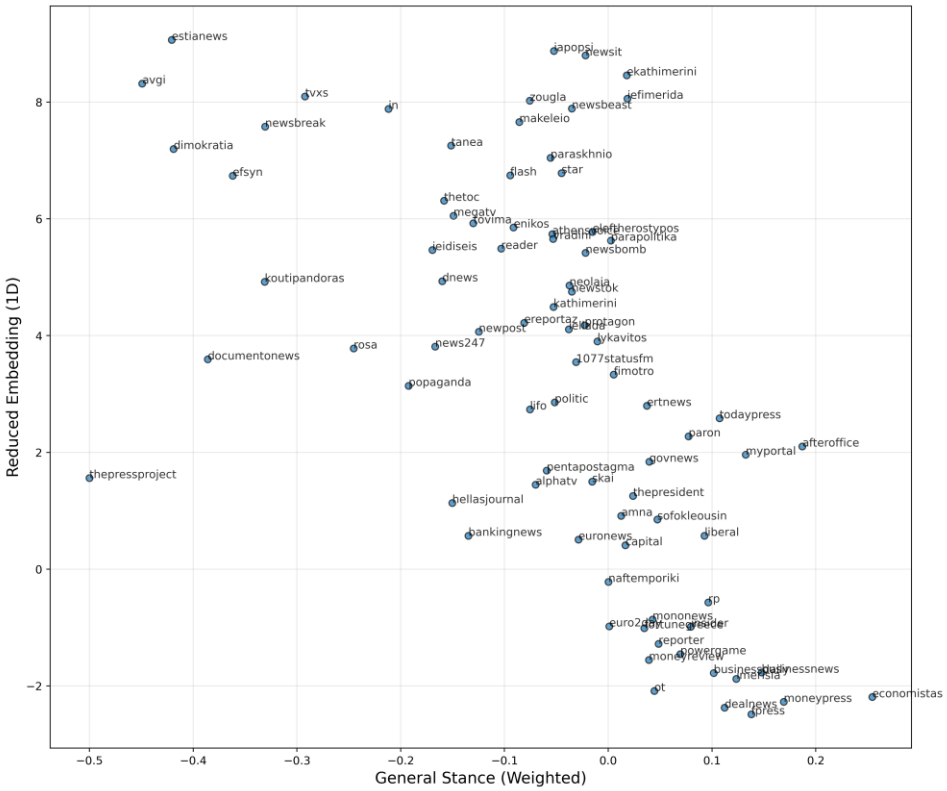
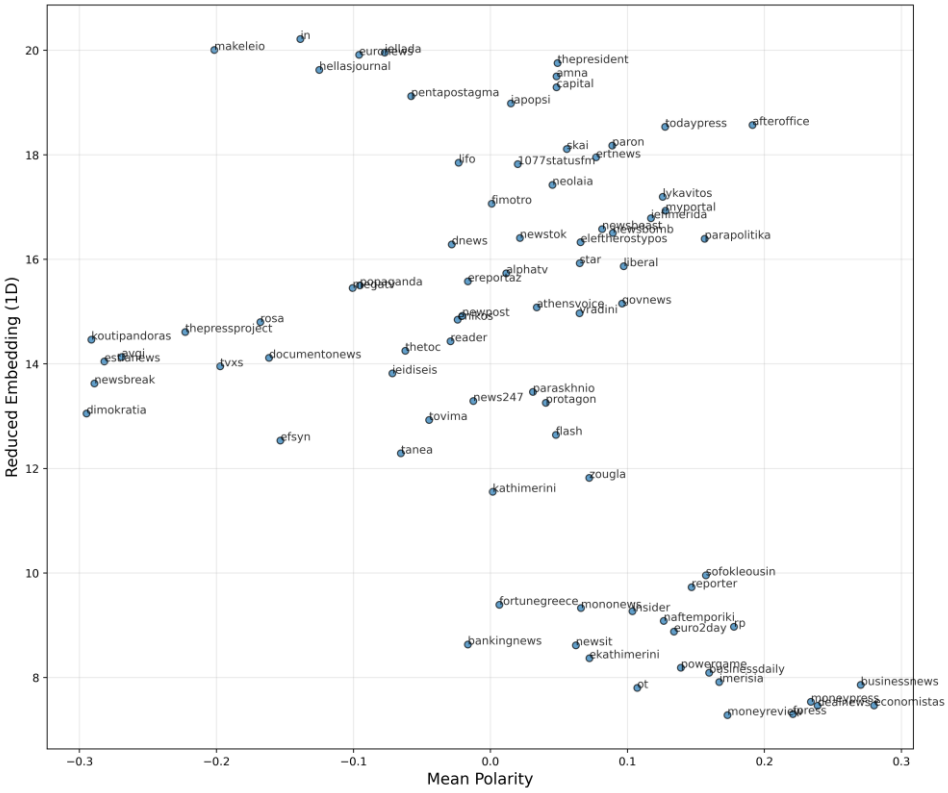


Figure 11. Mean Polarity vs Reduced Embedding (1D)



4.4 News Agency Content Reproduction

In this section, the study assesses the degree of content homogenization by means of news agency content diffusion within the Greek online media sphere. Often, newsrooms in Greek media fill their production with articles drawn from the Athens-Macedonian News Agency (AMNA - *Αθηναϊκό-Μακεδονικό Πρακτορείο Ειδήσεων, ΑΜΠΕ).* Although not necessarily something negative, normatively speaking, as newsrooms can be restricted by a variety of reasons (including lack of resources) such practices—where multiple outlets publish identical or near-identical stories—are common in media systems and can obscure the appearance of editorial independence or diversity.

The existing research examining news media copying of agency content spans multiple countries and methodological approaches. Studies have employed various methods including computational analysis, content analysis, and mixed methods approaches (e.g. Boumans et al., 2016). The extent of content copying varies significantly across media types and contexts.

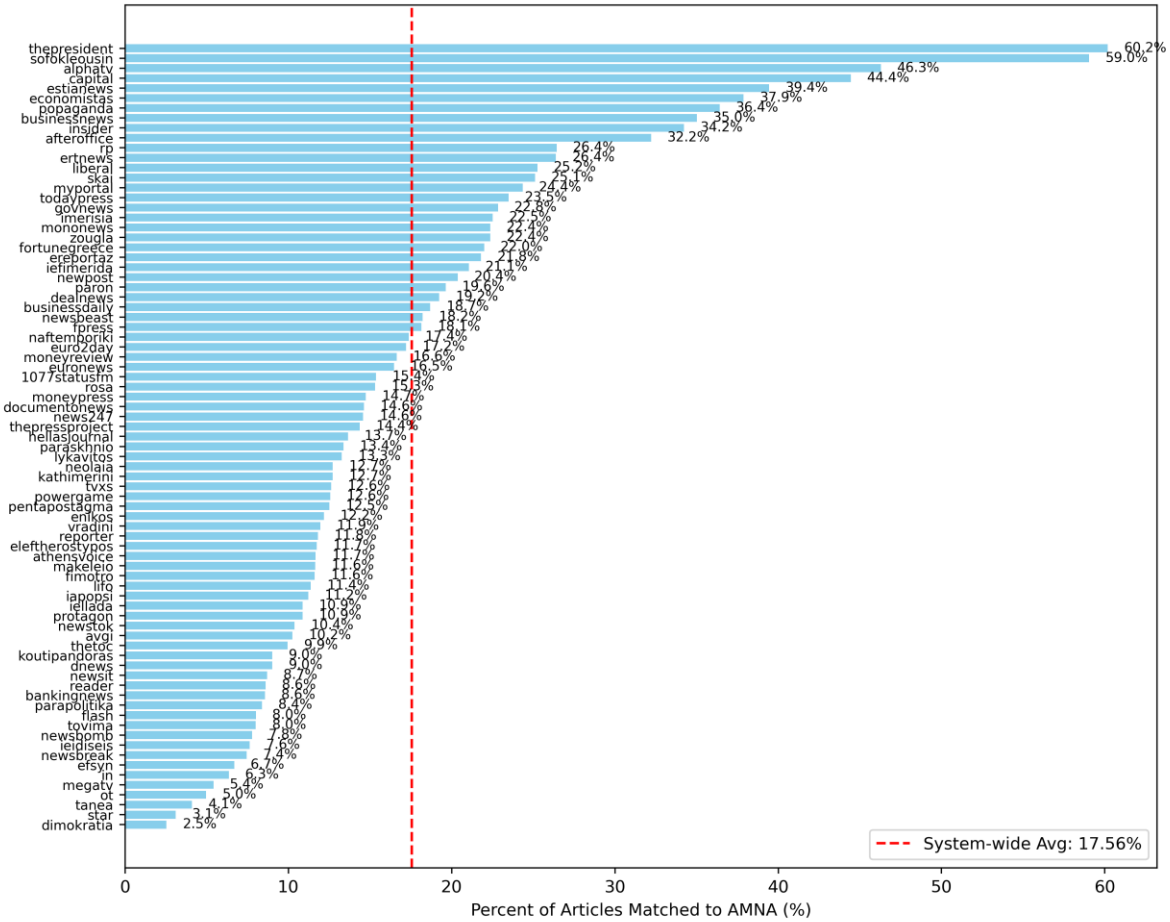
Data from news websites have not demonstrated a clear pattern with some research showing elevated reliance on agency content (Johnston and Forde, 2009), and some recent research indicating lower levels of content reuse than previously observed (Nicholls, 2019). Economic constraints and newsroom resource limitations consistently emerged as key factors driving increased reliance on agency content, particularly in digital formats. Studies across different geographical contexts, including the Netherlands, United Kingdom, and Australia, have documented this phenomenon, suggesting it represents a broader trend in contemporary journalism (Lewis et. al., 2008).

To detect potential news agency copy, all article titles originating from AMNA are used to construct a FAISS (Facebook AI Similarity Search) index based on the precomputed 384-dimensional sentence embeddings from the previous section. These embeddings are normalized, and cosine similarity is used to compare each non-AMNA title to the most similar AMNA title. Comparisons are restricted to articles published within the same ISO calendar week to account for temporal proximity in reproduction patterns. A similarity threshold of 0.9 is applied to identify strong matches, yielding a dataset of article pairs likely representing news agency content duplication.

This matching process allows for the calculation of duplication rates per outlet, visualized in

Figure 12, which shows the proportion of articles per website that closely resemble content from AMNA, along with the system-wide average. Additionally, Table A3 in the appendix contains a random sample of 10 matched article pairs to qualitatively illustrate patterns of content replication across different news sources. According to the data, and at this threshold level of similarity, the overall system-wide level of AMNA article reproduction stands at 17.56%[18]. However, there are multiple websites way beyond that average with some reproducing AMNA articles with rates as high as 60% (*thepresident*). Similar studies employing similar methods, find scores ranging from 4% to 11% (Garz & Ots, 2025).

Figure 12. Shares of Article Titles Likely Sourced from AMNA per Website



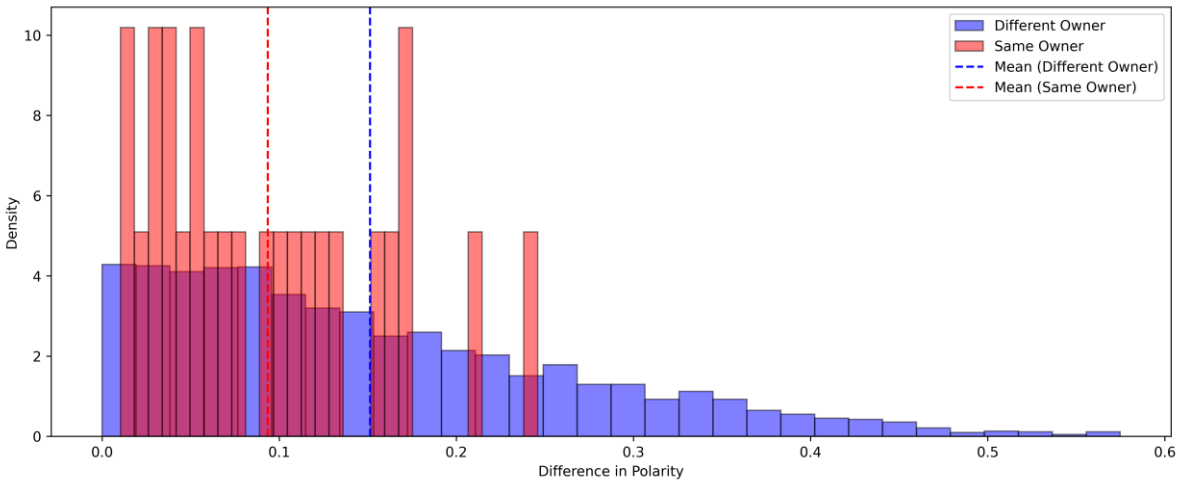
[18] The 0.9 threshold is a strict threshold that ensures high confidence in the similarity of article titles. Lowering this threshold yields much higher reproduction rates especially at the system level.

4.5 The Impact of Ownership

In this section, the study assesses the impact of ownership on content pluralism. To capture common ownership between websites, the study relies on company and financial data available on the register (Tax Identification Numbers - *ΑΦΜ*), previous studies and publications, investigative journalism efforts on the matter, as well as some manual expert annotation.[19]

Out of the 530 websites that are included in the original snapshot of the register, 51 (9.6%) belonged (at the time of writing) to one of the few powerful entrepreneurs mentioned in previous efforts in the literature and investigative journalism pieces (e.g. Ioannis Alafouzos, Thodoris Kyriakou, Evangelos Marinakis, Giannis Vardinogiannis, Dimitris Giannakopoulos, Dimitris Melisanidis, Ivan Savvidis, Ioannis Kaimenakis, and Dimitris Mpakos). These websites account for 29% of traffic in clicks[20] while at the same time received 38% of all funding from the "Stay at Home" Campaign (or "Petsas' list").[21] In the subset of 81 websites for which article data are collected and are included in the previous sections of the study, 19 websites (23.5%) belong to this group of entrepreneurs.

Figure 13. Distribution of Differences in Polarity between Websites

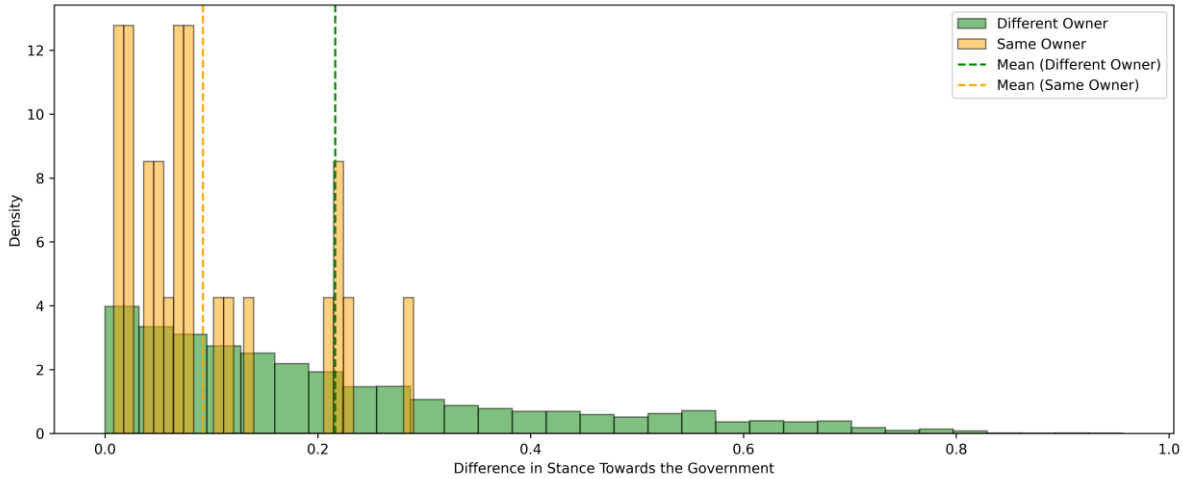


[19] This includes the Euromedia Ownership Monitor https://media-ownership.eu/, and Solomon's "Who Owns the Media" investigation project https://whoownsthemedia.gr/en. A complete list of website ownership links is available in Table A2 in the appendix.

[20] According to data drawn from semrush.com, a Search Engine Optimization (SEO) tool

[21] A fund for supporting the country's media industry due to financial stress induced by the Covid-19 pandemic. It was broadly and heavily criticised for the lack of transparency among other things (Souvlis, & Milonas, 2024). Journalistic investigation revealed that the outlets opposing the government received a disproportionally low share of the funds. See more at https://govwatch.gr/en/reports/greece-s-parliamentary-election-and-the-rule-of-law/

Figure 14. Distribution of Differences in Government Stance between Websites



All websites and the respective article titles are compared in a pairwise manner leading to 3240 unique pairs of websites. For each website pair the difference between the average scores of polarity and weighted stance towards the government (as described earlier) is calculated. Figures 13 and 14, plot these differences across all website pairs by whether they are under the same ownership or not. It is evident from this data, that the mean difference score when the websites are under the same ownership is much lower that when they are not, in both metrics.

Additionally, the differences between websites for all 3240 pairs (polarity and government stance, both overall and only for subsets of articles only about politics), are modeled using a standard OLS regression to try to see if the effect of common ownership holds statistically (Table 1). The data shows that common ownership reduces the differences between websites in terms of polarity and government stance both overall and in terms of political news only.

Table 1. Predicting Differences in Content (Polarity & Government Stance) between Websites – Standard OLS

	Polarity	**Government Stance**	**Polarity (Politics-only)**	**Government Stance (Politics-only)**
constant	0.1512***	0.2162***	0.1284***	0.2393***
	(-0.002)	(-0.0033)	(-0.0018)	(-0.0034)
Common Ownership	-0.0577**	-0.1243***	-0.0787***	-0.1383***
	(-0.0233)	(-0.0378)	(-0.0206)	(-0.039)
R-squared	0.0019	0.0033	0.0045	0.0039
R-squared Adj.	0.0016	0.003	0.0042	0.0036
N	3240	3240	3240	3240

Standard errors in parentheses.

* p<.1, ** p<.05, ***p<.01

5. Conclusions

Based on the empirical examination of the Greek online media sphere, this study sheds light on the apparent paradox of media (un)freedom. While Greece exhibits a numerically fragmented media landscape, particularly in the digital sphere, this quantitative fragmentation does not empirically translate into a genuine pluralism of content and viewpoints. The analysis of an exhaustive, large-scale dataset of over 3.37 million online news articles, from 81 websites, published in 2024, employing advanced NLP techniques such as sentiment analysis, stance detection, topic detection, and semantic embeddings, reveals seemingly low pluralism in terms of content and viewpoints.

The study found that at a system level, the Greek online media sphere is thematically non-divergent, meaning that websites tend to cover the same broad topics, indicating a relative

homogeneity in the news agenda. This suggests that despite the abundance of outlets, the diversity of information available to the public in terms of broad topics might not be as extensive as the fragmentation implies.

Furthermore, the analysis of news polarity and stance towards the government reveals significant patterns. Somewhat unexpectedly, a slight majority of Greek online news articles in the examined categories (Politics and Economy/Business) exhibit a positive polarity meaning that the overall system is slightly more positive in its news reporting. Perhaps even more unexpectedly, the system also does not seem to be considerably pro-government with a slight majority of articles being pro-government.

However, the study identified distinct clusters of websites based on their mean polarity and stance towards the government, with one larger cluster being somewhat negative but pro-government, and a smaller cluster being more markedly negative and anti-government. According to this data, not many websites report negatively on the government but those that do, do so more fiercely. Additionally, it is important to note here that a possible pro-government stance might be hidden under the cloak of criticism towards the opposition. It is possible that websites choose to support the government by means of criticizing the opposition, its parties or its leaders instead of directly praising the government. They often do so by showing opposition in a negative light, or casting doubt on the broader political landscape beyond the government.

Further potential limitations to editorial independence and viewpoint diversity in Greece are underscored by the observed level of news agency content reproduction, with a considerable portion of articles closely resembling content from the Athens-Macedonian News Agency (AMNA). While reliance on news agencies can be driven by resource constraints, the high duplication rates in some Greek online outlets may obscure the appearance of independent reporting and contribute to content homogenization. Similarly, the study's empirical clustering of websites when considering semantic embeddings, suggests underlying structural patterns in how news is framed and presented in Greece in terms of content. The analysis highlighted that political news tends to be discussed in a more similar manner across websites compared to other topics

Crucially, the study provides empirical evidence for the influence of media ownership on content. The evidence indicates that a relatively small number of powerful entrepreneurs own a substantial portion of the online media landscape. More importantly, the study found

a statistically significant correlation between common ownership and reduced differences in both polarity and government stance between websites. This means that websites that belong to the same owner tend to show more similar stances in political and economic news. These findings support concerns that media ownership structures in Greece exert a considerable influence on the diversity of viewpoints presented, potentially hindering true media pluralism despite the apparent fragmentation.

Finally, the clustering of websites based on sentiment and government stance, alongside the evidence of ownership influence and news agency content reproduction, collectively suggests that the apparent media fragmentation in Greece may mask a less diverse and potentially more concentrated information ecosystem in terms of actual content.

The implications of these findings for the quality of democracy in Greece are significant. A media landscape where quantitative fragmentation does not guarantee qualitative pluralism, and where ownership structures exert a discernible influence on content, can hinder informed public discourse and limit the range of perspectives available to citizens. The documented low trust in media in Greece further exacerbates these concerns, potentially impacting the media's role in holding power accountable and fostering a healthy democratic debate.

Future research should delve deeper into the mechanisms through which media ownership influences content production in the digital sphere, potentially employing qualitative case studies alongside large-scale quantitative analyses. Investigating the role of journalistic autonomy within different ownership structures and the impact of online advertising market dynamics on content diversity are also crucial avenues for further inquiry. Finally, the paper is also limited by its cross-sectional design that corresponds to only one country and only one government in it. Further comparative research should employ both within- and between-country designs that could offer valuable context and potential generalizable policy recommendations for enhancing media pluralism.

6. References

Angelou, I., Katsaras, V., Kourkouridis, D., & Veglis, A. Social Media Metrics as Predictors of Publishers' Website Traffic. Journalism and Media. 2024; 5(1):281-297. https://doi.org/10.3390/journalmedia5010019

Baker, C. E. (2006). *Media concentration and democracy: Why ownership matters*. Cambridge University Press.

Bashiri, H., & Naderi, H. (2024). Comprehensive review and comparative analysis of transformer models in sentiment analysis. *Knowledge and Information Systems*, *66*(12), 7305-7361.

Baum, M. A., & Zhukov, Y. M. (2019). Media ownership and news coverage of international conflict. *Political Communication*, *36*(1), 36-63.

Blair, S. J., Bi, Y., & Mulvenna, M. D. (2020). Aggregated topic models for increasing social media topic coherence. *Applied intelligence*, *50*, 138-156.

Bleyer-Simon, K., Da Costa Leite Borges, D., Brogi, E., Carlini, R., Kermer, J. E., Palmer, M., ... & Žuffová, M. (2024). Monitoring media pluralism in the digital era: application of the media pluralism monitor in the European member states and in candidate countries in 2023. European University Institute.

Boumans, J. W., Vliegenthart, R., & Boomgaarden, H. G. (2016). Nuclear voices in the news: A comparison of source, news agency and newspaper content about nuclear energy over time. *European Journal of Communication*, *31*(3), 260-282.

Champion, K. (2015). Measuring content diversity in a multi-platform context. *The Political Economy of Communication*, *3*(1).

Costola, M., Hinz, O., Nofer, M., & Pelizzon, L. (2023). Machine learning sentiment analysis, COVID-19 news and stock market reactions. *Research in International Business and Finance*, *64*, 101881.

Garz, M., & Ots, M. (2025). Media consolidation and news content quality. *Journal of Communication*, jqae053.

Garz, M., Ots, M., & Sjøvaag, H. (2024). Political viewpoint diversity in the news: Market and ownership conditions for a pluralistic media system. *The International Journal of Press/Politics*, *29*(4), 983-1003.

Hanretty, C. (2014). Media outlets and their moguls: Why concentrated individual or family

ownership is bad for editorial independence. *European Journal of Communication*, *29*(3), 335-350.

Hendrickx, J., & Ranaivoson, H. (2021). Why and how higher media concentration equals lower news diversity–The Mediahuis case. Journalism, 22(11), 2800-2815.

Hendrickx, J., & Kammer, A. (2024). Transnational chain ownership of news media: Integration and independence across borders in DPG media. *International Communication Gazette*, 17480485241288769.

Hendrickx, J., & Van Remoortere, A. (2024). Newspaper ownership, democracy and news diversity: a quantitative content homogeneity study. *Journalism Practice*, *18*(10), 2531-2546.

Johnston, J., & Forde, S. (2009). 'Not wrong for long': the role and penetration of news wire agencies in the 24/7 news landscape. *Global Media Journal*, *3*(2), 1-16.

Koch, A., Huynh, T. L. D., & Wang, M. (2024). News sentiment and international equity markets during BREXIT period: A textual and connectedness analysis. *International Journal of Finance & Economics*, *29*(1), 5-34.

Korbiel, I., & Sarikakis, K. (2017). Between two (in) competencies: A critical view on media ownership regulation in the EU. *International Journal of Media & Cultural Politics*, *13*.

Lewis, J., Williams, A., & Franklin, B. (2008). A compromised fourth estate? UK news journalism, public relations and news sources. *Journalism studies*, *9*(1), 1-20.

Newman, N., Fletcher, R., Robertson, C. T., Ross Arguedas, A., & Nielsen, R. K. (2024). Reuters Institute digital news report 2024. Reuters Institute for the Study of Journalism.

Nicholls, T. (2019). Detecting textual reuse in news stories, at scale. *International Journal of Communication*, *13*(2019).

Noam, E. M. (Ed.). (2016). *Who owns the world's media?: media concentration and ownership around the world*. Oxford University Press.

Papadopoulou, L. (2020). Alternative hybrid media in Greece: An analysis through the prism of political economy. *Journal of Greek Media & Culture*, *6*(2), 199-218.

Papadopoulou, L., Angelou, I. (2024), Monitoring media pluralism in the digital era : application of the media pluralism monitor in the European member states and in candidate countries in 2023. Country report : Greece, European University Institute - https://hdl.handle.net/1814/77003

Papathanassopoulos, S., Karadimitriou, A., Kostopoulos, C., & Archontaki, I. (2021). Greece: Media concentration and independent journalism between austerity and digital disruption. In J. Trappel, & T. Tomaz (Eds.), The Media for Democracy Monitor 2021: How leading news media survive digital transformation (Vol. 2) (pp. 177–230). Nordicom, University of Gothenburg. https://doi.org/10.48335/9789188855428-5

Pinto, S., Albanese, F., Dorso, C. O., & Balenzuela, P. (2019). Quantifying time-dependent media agenda and public opinion by topic modeling. *Physica A: Statistical Mechanics and Its Applications*, *524*, 614-624.

Rozado, D., Hughes, R., & Halberstadt, J. (2022). Longitudinal analysis of sentiment and emotion in news media headlines using automated labelling with Transformer language models. *Plos one*, *17*(10), e0276367.

Sjøvaag, H., & Ohlsson, J. (2019). Media ownership and journalism. In *Oxford research encyclopedia of communication*.

Souvlis, G., & Milonas, P. P. (2024). State and Nationalism: The Greek Media's Coverage of the COVID-19 Pandemic. In *The Nation in the Time of the Pandemic: Media and Political Discourse across Countries during the COVID-19 Crisis* (pp. 251-268). Cham: Springer International Publishing.

Trappel, J., & Meier, W. A. (2022). Soaring media ownership concentration: Comparing the effects of digitalisation on media pluralism and diversity.

Vogler, D., Udris, L., & Eisenegger, M. (2020). Measuring media content concentration at a large scale using automated text comparisons. *Journalism Studies*, *21*(11), 1459-1478.

Appendix

Table A1. IPTC Media Topics Taxonomy and Descriptions

Category Name	*Category Description*
arts, culture, entertainment and media	All forms of arts, entertainment, cultural heritage and media
conflict, war and peace	Acts of socially or politically motivated protest or violence, military activities, geopolitical conflicts, as well as resolution efforts
crime, law and justice	The establishment and/or statement of the rules of behaviour in society, the enforcement of these rules, breaches of the rules, the punishment of offenders and the organisations and bodies involved in these activities
disaster, accident and emergency incident	Man made or natural event resulting in loss of life or injury to living creatures and/or damage to inanimate objects or property
economy, business and finance	All matters concerning the planning, production and exchange of wealth.
education	All aspects of furthering knowledge, formally or informally
environment	All aspects of protection, damage, and condition of the ecosystem of the planet earth and its surroundings.
health	All aspects of physical and mental well-being
human interest	Item that discusses individuals, groups, animals, plants or other objects in an emotional way
labour	Social aspects, organisations, rules and conditions affecting the employment of human effort for the generation of wealth or provision of services and the economic support of the unemployed.
lifestyle and leisure	Activities undertaken for pleasure, relaxation or recreation outside paid employment, including eating

	and travel.
politics	Local, regional, national and international exercise of power, or struggle for power, and the relationships between governing bodies and states.
religion	Belief systems, institutions and people who provide moral guidance to followers
science and technology	All aspects pertaining to human understanding of, as well as methodical study and research of natural, formal and social sciences, such as astronomy, linguistics or economics
society	The concerns, issues, affairs and institutions relevant to human social interactions, problems and welfare, such as poverty, human rights and family planning
sport	Competitive activity or skill that involves physical and/or mental effort and organisations and bodies involved in these activities
weather	The study, prediction and reporting of meteorological phenomena

Table A2. Websites Collected and their Ownership Links

Website	Ownership Big9	MatchingVAT	MHT VAT
1077statusfm.gr			82500851
afteroffice.gr		61	76753850
alphatv.gr	Vardinogiannis		94438520
amna.gr			998106073
athensvoice.gr		370	999600491
avgi.gr			94445480
bankingnews.gr			800834856
businessdaily.gr			801188759
businessnews.gr		194	800497000
capital.gr			998771128
dealnews.gr			802375888
dimokratia.gr			800644028
dnews.gr			800961697
documentonews.gr			800808181
economistas.gr			801043626
efsyn.gr			800428018
ekathimerini.com	Alafouzos	316	996589046
eleftherostypos.gr			801129700
enikos.gr		184	800384700
ereportaz.gr			800545801
ertnews.gr			Not on MHT
estianews.gr			92178470
euro2day.gr		108	94542790
euronews.com			Not on MHT
fimotro.gr			45783311
flash.gr	Mpakos-		99553793

	Kaimenakis		
fortunegreece.com			801432933
fpress.gr			800712933
govnews.gr		357	998606750
hellasjournal.com			Not on MHT
iapopsi.gr			800706834
iefimerida.gr		350	998381058
ieidiseis.gr			801070002
iellada.gr	Giannakopoulos		Not on MHT
imerisia.gr	Savvidis	88	94249653
in.gr	Marinakis	222	800745939
insider.gr		203	800565178
kathimerini.gr	Alafouzos	316	996589046
koutipandoras.gr			998408813
liberal.gr			800862136
lifo.gr			998936241
lykavitos.gr		206	800587042
makeleio.gr			801962016
megatv.com	Marinakis	222	800745939
moneypress.gr		206	800587042
moneyreview.gr	Alafouzos	316	996589046
mononews.gr			800866100
myportal.gr		265	801343660
naftemporiki.gr	Melissanidis	320	996975246
neolaia.gr	Alafouzos		Not on MHT
newpost.gr	Vardinogiannis		800689593
news247.gr		125	99456944
newsbeast.gr		332	997801220

newsbomb.gr	Giannakopoulos	326	997434600
newsbreak.gr			800812606
newsit.gr			997837935
newstok.gr		175	158245168
ot.gr	Marinakis		801010853
**parapolitika.gr*	Marinakis	207	800595750
paraskhnio.gr		249	801016102
paron.gr		336	997883048
pentapostagma.gr			Not on MHT
popaganda.gr			800424148
powergame.gr		271	801489648
protagon.gr			997984911
reader.gr		203	800565178
reporter.gr			94436201
rosa.gr			801776035
rp.gr			800501496
skai.gr	Alafouzos	363	998921252
sofokleousin.gr			800883918
star.gr			94281721
tanea.gr	Marinakis	222	800745939
thepresident.gr			800949819
thepressproject.gr			Not on MHT
thetoc.gr	Filippopoulos	106	94460077
todaypress.gr		61	76753850
tovima.gr	Marinakis	222	800745939
tvxs.gr			996788127
vradini.gr			94457860
zougla.gr		351	998382123

**Only back to 2/11/2024*

Table A3. Examples of Article Titles from AMNA and their Matches

	Title	Source	Matched AMNA Title	Similarity
475720	Υπερταμείο/Growthfund: Δημιουργία αξίας 0,7 δισ. ευρώ το διάστημα 2022-2024	myportal	Υπερταμείο/Growthfund: Δημιουργία αξίας 0,7 δισ. ευρώ το διάστημα 2022-2024 - Στον τελευταίο 1,5 χρόνο πέτυχε περισσότερα κέρδη από το σύνολο όλων των προηγούμενων χρήσεων	0.934
494928	Βοσνία-Ερζεγοβίνη: Συνελήφθη ο υπουργός Ασφαλείας κατηγορούμενος για διαφθορά	news247	Βοσνία-Ερζεγοβίνη: Συνελήφθη ο υπουργός Ασφαλείας κατηγορούμενος για διαφθορά	1.000
43032	ΔΥΠΑ: Στις 7/2 η πρώτη διαδικτυακή ευρωπαϊκή «Ημέρα Καριέρας Τουρισμού»	economistas	Ολοκληρώθηκε η πρώτη διαδικτυακή ευρωπαϊκή «Ημέρα Καριέρας» Τουρισμού της ΔΥΠΑ	0.964
19200	Κρίσεις Αστυνομικών Διευθυντών της Ελληνικής Αστυνομίας – Ποιοι προάγονται στον βαθμό του ταξίαρχου	newsbeast	Κρίσεις Αστυνομικών Διευθυντών της Ελληνικής Αστυνομίας	0.970
482289	Ερώτηση Μ. Κατρίνη: Να μπει φραγμός στην εξαγωγή ευρωπαϊκών οπλικών συστημάτων στην Τουρκία	efsyn	Βουλή Ερώτηση Μ. Κατρίνη: Να μπει φραγμός στην εξαγωγή ευρωπαϊκών οπλικών συστημάτων στην Τουρκία	0.990
53275	Υπό συζήτηση παύση των	govnews	Μεσανατολικό: Συζητείται	0.953

	εχθροπραξιών για τουλάχιστον 6 εβδομάδες στη Λωρίδα της Γάζας		«παύση» των εχθροπραξιών για «τουλάχιστον 6 εβδομάδες» στη Λωρίδα της Γάζας (πρόεδρος των ΗΠΑ)	
367222	Χαλκιδική: Εξηγήσεις καλούνται να δώσουν τρεις για το δυστύχημα στο λούνα παρκ	capital	Χαλκιδική: Εξηγήσεις καλούνται να δώσουν τρεις για το δυστύχημα στο λούνα παρκ	1.000
41977	Βασιλικός (ΞΕΕ): Γνώση και συνεργασία η «συνταγή» απέναντι στις μεγάλες προκλήσεις του τουρισμού	businessnews	Αλεξ. Βασιλικός: Γνώση και συνεργασία η «συνταγή» απέναντι στις μεγάλες προκλήσεις του τουρισμού	0.959
298103	ΑΑΔΕ: Αποκλειστικά ψηφιακά από 1/9 η δέσμευση και αποδέσμευση του αποδεικτικού ενημερότητας για χρέη σε δήμους και περιφέρειες	thepresident	ΑΑΔΕ: Αποκλειστικά ψηφιακά από 1/9 η δέσμευση και αποδέσμευση του αποδεικτικού ενημερότητας για χρέη σε δήμους και περιφέρειες	1.000
42944	Σκρέκας: Στρατηγική αυτονομία της Ε.Ε. με ενιαία βιομηχανική πολιτική	economistas	Κ. Σκρέκας: Στρατηγική αυτονομία της ΕΕ με ενιαία βιομηχανική πολιτική	0.914

0.970169