9-2017

Convergence of media attention across 129 countries

Jisun AN

Hassan ALDARBESTI

Haewoon KWAK
Singapore Management University, hkwak@smu.edu.sg

Convergence of Media Attention Across 129 Countries

Jisun An(✉), Hassan Aldarbesti, and Haewoon Kwak

Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar
{jisun.an,haewoon}@acm.org, aldarbesti@outlook.com

Abstract. The objective of this study is to assess the longitudinal trends of media similarity and dissimilarity on the international scale. As news value has well-established political, cultural, and economic consequences, the degree to which media coverage and content is converging across countries has implications for international relations. To study this convergence, we use the daily data of the 100 topics that were over-reported in each country, compared to other countries, from March 7 to October 9, 2016. The results of this analysis indicate that two complementary patterns–globalization and domestication–explain the media attention across the countries. We conclude that this attention can be driven not only by geographical closeness but also by more complex dimensions, such as historical relationships. Also, although a group of countries often have common media attention, their similarity level depends on time and topic.

Keywords: Media convergence · Media attention · Globalization · Domestication · Unfiltered News · Tensor factorization

1 Introduction

Media plays a significant role in determining what events are important and providing information to audiences on those events [11]. In the era of technological advancement and globalization, there is a growing interest in the impact of these factors on the media–on media globalization and convergence [8]. While early studies of these concepts focused on the systematic flow of media from international sources to advance our understanding of the factors that influence domestic coverage of foreign events [9], globalization and technology have increased the flow of news between countries, as well as our access to that information. There is a long tradition in media studies looking at the systemic determinants of news content and coverage of international news topics and events [17], but there is a lack of empirical evidence on how much the manner in which the media presents events across countries is converging.

As "transnational information flow is a reflection and a constituent of the larger global system" [16], understanding the similarity of media values and content across countries has significant political, economic, and cultural implications. To this effect, this study evaluates the over and under-emphasis of news

G.L. Ciampaglia et al. (Eds.): SocInfo 2017, Part II, LNCS 10540, pp. 159–168, 2017.
DOI: 10.1007/978-3-319-67256-4_14

topics over time and determines how this compares by country in order to assess if there is, in fact, a pattern of convergence at the international scale. We assess the homogenization of media attention over time using Unfiltered News and 100 media topics and hypothesize that, through globalization forces, media content is, in fact, converging.

2 Related Work

The gap in research that this study fills has been identified through a review of the extant literature on media content, convergence, and scale. The dominant theme is the literature on media attention and similarity on an international level, which is the comparison of multiple countries to assess how a selected event or topic has been covered (c.f., [16]). Media convergence occurs when sources become more similar. Regarding scale, it is important to understand if this convergence is happening at the domestic level, at the international level, or both. The structures of media content and attention are assessed through this review of the extant literature. News convergence is defined as an increase in the similarity of content over time and scale (i.e., domestic and international).

With increased attention to both media content and globalization, there is growing concern in the academic community that media content is converging. One side of the debate hypothesizes that agency copy within the media [4] results in cultural imperialism by which US news, culture, and values are shared and adopted international through media [7]. However, there is another side of literature that presents empirical evidence challenges this claim. In fact, Nielsen (2013) presents enduring differences and divergences across media. Flew and Waisbord (2015) makes a similar argument for national media politics [8]. News coverage and convergence of this coverage is of interest for their meaning as "news values." These are values that give meaning to modern society. As there are millions of events that occur everyday in the world, news values dictate those stories that are most salient [10]. The question remains as to whether these news values are converging.

While there are arguments both for and against news convergence, no studies to date have measured the direct changes in the homogenization of news coverage over time. Although conducting a quantitative analysis of how media covered the Israeli-Palestinian conflict over a decade [3], the research presented herein covers a wide range of topics. We note that Baden and Tenenboim-Weinblatt (2016) conclude that cultural and language affinity is the underlying factor of media content selection patterns, but that these findings only relate to the conflict of study and, as so, cannot be generalized to speak to media convergence more broadly [3]. Our study, however, presents a more generalizable assessment of news convergence through looking at the similarity of media content more broadly over time.

Most of the studies have conducted with limitations of studying a handful number of countries, news articles, or topics. Privileged in the digital era and

the emergence of advanced machine learning techniques, we overcome all these limitations and provide a data-driven analysis of media attention of 196 countries during 211 days.

3 Data Collection

Unfiltered News offers two kinds of indexed data for each country: topics mentioned more than other topics (In the rest of the paper, we simply say "topics mentioned more") and topics mentioned less than other countries do (we say "topics mentioned less"). The former represents what media pay attention, and the latter reflects what media do not pay attention. In this work, we focus only on the topics mentioned more that encode media attention. A deeper analysis of media attention (topics mentioned more) and media disregard (topics mentioned less) together is available in [12]. We collect the data from 7 March to 9 October 2016.

3.1 Data Description and Notation

Unfiltered News ideally provides the 100 topics mentioned more for each country; however, it sometimes has missing data. For a fair comparison across the countries, we filter out such incomplete data.

We denote by $\mathrm{C} = \{c_1, c_2, ..., c_n\}$ a set of n countries and $\mathrm{D} = \{d_1, d_2, ..., d_m\}$ a set of m days available in our data. We denote by $\mathrm{M}_{d_j}^{c_i}(k)$ a set of k topics mentioned more in the country c_i on the day d_j. For example, $\mathrm{M}_{20160101}^{Korea}(10)$ is a set of the top 10 most frequently mentioned topics in Korea on 1 January 2016. We then refine the data in the following ways:

1. We set the threshold k $(0 < \mathrm{k} \leq 100)$.
2. For each country $c_i \in \mathrm{C}$, we check whether at least k topics mentioned more are available on $\forall d_i \in \mathrm{D}$. If so, we add c_i to a set C^k.
3. Let us say $\mathrm{C}^k = \{c_1^k, c_2^k, ..., c_l^k\}$.
4. Then, the final datasets of the topics mentioned more with k, $\mathrm{M}(k)$, is:
   $\mathrm{M}(k) = \{(\mathrm{M}_{d_1}^{c_1^k}(k), \mathrm{M}_{d_2}^{c_1^k}(k), ..., \mathrm{M}_{d_m}^{c_1^k}(k)), \ (\mathrm{M}_{d_1}^{c_2^k}(k), \ \mathrm{M}_{d_2}^{c_2^k}(k), \ ..., \ \mathrm{M}_{d_m}^{c_2^k}(k)), \ ..., \ (\mathrm{M}_{d_1}^{c_l^k}(k), \mathrm{M}_{d_2}^{c_l^k}(k), ..., \mathrm{M}_{d_m}^{c_l^k}(k))\}$.

When we do not have any constraint $(k = 0)$, the number of countries in C^0 is 196. With the weakest condition $(k = 10)$, $|\mathrm{C}^{10}|$ quickly decreases to 129. The number of countries in C^k monotonically decreases with growing k and reaches at 88 when $k = 90$. We conducted all our experiments with different k and found that the overall trend stays the same with some variations in numbers. Thus, in this work, we report the result with specific k (usually 10) and omit other results due to lack of space.

4 Similarity and Dissimilarity of Media Attention Across the Countries

4.1 Globalization and Domestication of Media

As we mentioned in the Related work section, there are several reasons why the media attention of different countries is expected to converge: (1) News published on the Web can have a worldwide audience, overcoming geographical limitations; thus, the media chooses news topics that appeal to as many people as possible; and (2) The role of a global news agency becomes more and more important. For financial reasons, it is impossible to have offices all over the world. Instead, the media relies on global news agencies, such as Reuters or AP, for foreign news. The fact that news coverage of foreign disasters is heavily dependent on whether it is reported by global news agencies [11] shows this trend.

By comparing the topics mentioned more often in each of the countries, we measure the similarities in media attention among the countries. We begin with how many countries mention the same topic every day.

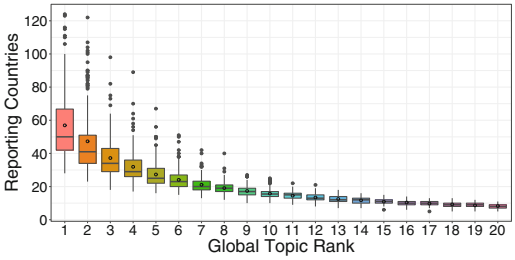

Fig. 1. How many countries report the globally popular topics

For each of 211 days, we count how many countries report each of the top global topics on the same day. Figure 1 shows how many countries mentioned the global top topics on a certain day, where $k = 10$. The variance becomes small, and the number of countries reporting the global topic decreases. The topics ranked first and second received media attention from 57 and 47 countries (out of 129 countries) as a median, respectively. The fact that tens of the countries share the top global topics every day shows a high level of globalization in the world.

As we manually look into which topic is the most widely mentioned in the world every day, we find some common topical characteristics: sports (e.g., Rio Olympics or UEFA Champions League), terror in the Western world (e.g., France, Germany, or Belgium), big natural disasters (e.g., earthquakes in Italy), airplane crash (e.g., EgyptAir crash), and big political events (e.g., coup in Turkey).

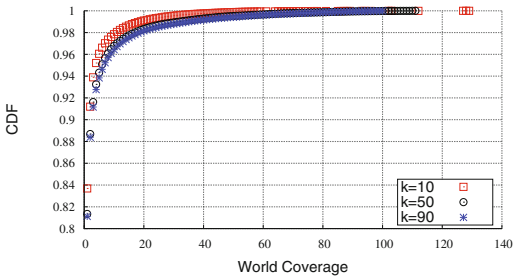

Fig. 2. Cumulative distribution of the number of countries that mention the same topic

In Fig. 2, by contrast, the rest of the media attention is highly unique. For clarity, we define the attention diversity as the union of the topics more mentioned and denoted it by $\mathrm{U}(k) = \bigcup_{c_j}^{\mathrm{C}^k} \bigcup_{d_i}^{D} \mathrm{M}_{d_i}^{c_j}(k)$ as in [1]. Then, we find that 83.7% of all the topics in U(10) appears in a single country. When testing with different ks, the proportion of the topics that get attention from a single country stays high (81.1% in U(90)). This uniqueness of media attention in each country demonstrates that media continues shaping news for domestic readers [3].

In summary, two complementary patterns–globalization and domestication–explain the media attention across the countries.

4.2 Countries Sharing Media Attention in a Long-Term

We begin with vectorizing media attention. We create a $|\mathrm{U}(k)|$-d vector space, $\mathbf{V}(k)$, where each dimension is mapping into each topic $t \in \{t_1, t_2, ..., t_{|\mathrm{U}(k)|}\}$. In $\mathbf{V}(k)$, we can locate a country c_i as a vector whose n-th element is the number of days ($|\forall d_j|$) that the topic $t_n \in \mathrm{M}_{d_j}^{c_i}(k)$. We build a distance matrix among country vectors and apply a hierarchical clustering method using Ward's method [14].

Figure 3 is the dendrogram to show the attention similarity among countries. In the figure, we clearly see the regional groups, which are consistent with our intuition: the Eastern Europe (at 12 o'clock), the Western and Northern Europe (at 2 o'clock), the MENA region (at 4 o'clock), Latin America (at 6 o'clock), and Asia and North America (at 9 o'clock). Seeing the leaf level, which is the direct connection between two countries, we find that most of them are neighboring countries, such as Cyprus and Greece, Belgium and France, Austria and Germany, Qatar and Kuwait, Ireland and the United Kingdom, the United States and Canada, and so on.

Some historical colonial ties are also closely located in the dendrogram. Even though Spain and Mexico are far apart geographically, they directly connect to each other on the tree. Singapore and Malaysia connect to the United Kingdom at the lower level rather than other Asian countries. Similarly, Philippines connects to the tree that contains the United States. These colonial ties do not only

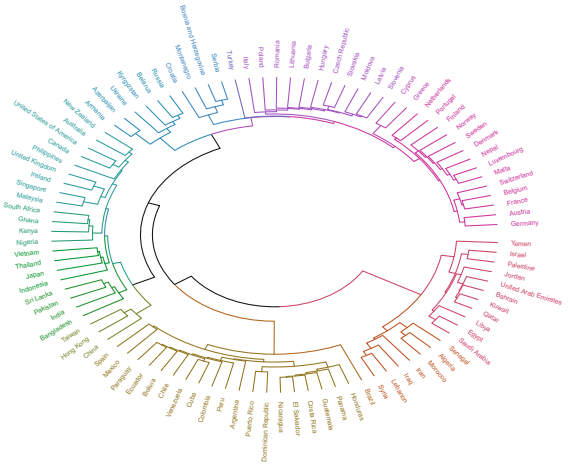

Fig. 3. [Zoomable in PDF] Attention similarity between countries regarding the topics more mentioned

remain in past relationships but have also been developed as strong economic and political relationships. Therefore, it is understandable that the news media in these countries share similar attention.

We can also see a block-level association in the dendrogram. There are several regional clusters in Europe (Northern, Eastern, Southern, and Western), and they are more similar to each other than other regions. Europe connects to North America, Asia, Latin America, and the MENA region. In other words, the MENA region is the most different from the rest of the world regarding the topics more mentioned by local media.

4.3 Country Blocks Having Common Media Attention

We found that there are a few topics that attract media attention from multiple countries. Are there any patterns in these "multiple" countries? To answer this question, we must systematically identify country blocks that have common media attention.

Instead of using an aggregated data, we build a 3-way tensor, $\mathcal{T}(k) \in \mathbb{R}^{N \times N \times S}$, where N is the number of countries in C^k, and S is the number of days in D, to represent the attention similarity among countries for each day; τ_{ijk} is the attention similarity between country c_i and c_j on day d_k, defined as cosine similarity between $\mathrm{M}^{c_i}_{d_k}(k)$ and $\mathrm{M}^{c_j}_{d_k}(k)$. This allows us to incorporate the timing of the media attention in measuring the similarity between countries.

By representing a tensor as a product of lower-dimensional factors, we can uncover its latent structure. Out of several techniques, we use PARAFAC decomposition for its simplicity and comparable performance [5]. A detailed discussion of the advantages and disadvantages of PARAFAC, versus its alternatives, such as Tucker3 and two-way PCA, is available in [5]. We also apply a constraint, with

the non-negativity of elements, in the resulting matrices for ease of interpretation. We use N-way Toolbox 3.31 for MATLAB [2] for the actual computation.

The PARAFAC decomposition of a 3-way tensor, $\mathcal{T} \in \mathbb{R}^{N \times N \times S}$, results in three matrices, **A**, **B**, and **C**, whose dimensions are $N \times R$, $N \times R$, and $S \times R$, respectively, and $R \ll min(rank(\mathcal{T})) = min(N, N, S)$ to benefit from the decomposition. Then, an element of $\mathcal{T}$ can be written as following:

$$\tau_{ijk} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} + \epsilon_{ijk} \tag{1}$$

In Eq. (1), R is the number of components that encode the level of detail. A higher R gets a more detailed latent structure but has the risk of over-fitting, and a lower R overlooks the latent structure but is resilient to noise. We use the core consistency that is proposed in [6], to systematically determine the best R. We run the PARAFAC model from 2 to 10 and compute the core consistency. By the scree plot and the rule of thumb, that core consistency higher than 0.5 is generally acceptable, we choose $R = 5$ (core consistency = 0.560). In other words, by the PARAFAC decomposition of $\mathcal{T}$, we get the five components from **A** and **B**. As the attention similarity matrix among countries is symmetric, **A** is the same as **B**.

Table 1. Top 10 countries (with the highest loading factors) in each component

CP	Countries
1	France, Belgium, Switzerland, Luxembourg, Austria, Germany, Slovakia, Czech Republic, Slovenia, Denmark
2	Saudi Arabia, Egypt, Kuwait, United Arab Emirates, Palestine, Yemen, Jordan, Qatar, Bahrain, Libya
3	Russia, Belarus, Ukraine, Kyrgyzstan, Azerbaijan, Moldova, Latvia, Bulgaria, Lebanon, Iran
4	Australia, United States of America, Singapore, South Africa, United Kingdom, Malaysia, New Zealand, Pakistan, China, Thailand
5	Venezuela, Colombia, Chile, Ecuador, Peru, Panama, Bolivia, Nicaragua, Costa Rica, Argentina

The resulting components are presented in Table 1. Interestingly, with the exception of one component (CP4), geographically close countries are well grouped together, such as the Western European countries (CP1), the Middle East and North African (MENA) countries (CP2), the Eastern European countries (CP3), and the Latin American countries (CP5). Since geographical proximity is known to be associated with culture, language, and ethnicity affinity, and also lead to strong economic and political relations, it is reasonable that countries in the same regional block share media attention. The CP4, which has both neighboring countries and distant countries together, is an interesting

mixture of geographical proximity (Asian countries) and colonial ties with the United Kingdom (the former British Empire). The result shows that the similar media attention of countries can be driven not only by geographical closeness but also by more complex dimensions, such as historical relationships.

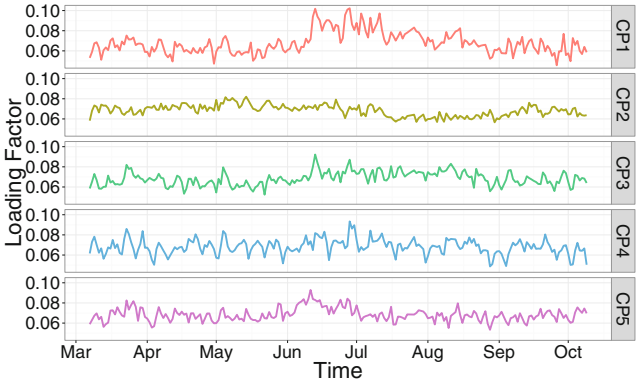

Fig. 4. Activate patterns of identified components

The remaining matrix, **C**, encodes the activate patterns of the identified components (i.e., when do countries in each component become more similar to each other?). Figure 4 shows constant fluctuations in the activation of each component. It means that, within each of components, the similarity of media attention among countries continuously changes. For example, the media attention of countries in CP1 become more similar around June and July, during the Brexit referendum, compared to other months. In other words, although the component we found is a group of countries who often have common media attention, their similarity level depends on time and topic.

5 Conclusion

Media is consistent in its roles of socialization and communication [11] but constantly evolving in in its technological basis and its impact [8]. Globalization, technological advancements, and a changing understanding of the role that media plays in the presentation of and access to information have warranted this study on media similarity and dissimilarity on an international scale.

Thus far, scholars have focused on understanding the current global divides and power structures through news coverage [15,16]. In understanding the factors relating to foreign news coverage based on limited data, studies have shown inconsistent results because of cultural, regional, or political differences [13].

While the determinants of news coverage have been studied at an international level, this work is the first empirical study of the convergence of media

coverage of events globally. We have evaluated the over-emphasis of news topics by country to assess whether there is a convergence of topic coverage at the international level by evaluating the longitudinal homogenization of media attention. Using Unfiltered News and 100 topics, we began with the hypothesis that globalization is, in fact, resulting in the convergence of news coverage across countries. Our results confirm the convergence of news coverage on a global scale, but we find that this is not the only pattern of convergence. In addition to a pattern of globalization, we find that coverage is also converging at the country level, as a form of domestication. As we estimate the press freedom index of a country on the basis of the topic diversity of news media in that country in our previous work [1], it would be possible to make the "globalization index" or "domestication index" in terms of the topic overlap among countries on the basis of what we did in this work. We will leave this for future research directions. Interestingly, we also find similarity patterns in the similarity across countries that can be traced to historical colonial ties. For instance, the dendrogram depicts similarity and shows direct links between Spain and Mexico; it also links Malaysia and Singapore to the United Kingdom and the Philippines to the United Kingdom. This demonstrates that the economic and political legacy of colonialism is evident in the current media coverage across countries. Finally, we find geographical proximity in convergence. At the block-level association in the dendrogram, we see regional clusters, such as European countries, which exhibit similarity.

References

1. An, J., Kwak, H.: Data-driven approach to measuring the level of press freedom using media attention diversity from Unfiltered News. In: The Workshops of the Eleventh International AAAI Conference on Web and Social Media AAAI Technical Report WS-17-17: News and Public Opinion (NECO), pp. 739–742 (2017)
2. Andersson, C.A., Bro, R.: The N-way toolbox for MATLAB. Chemometr. Intell. Lab. Syst. **52**(1), 1–4 (2000)
3. Baden, C., Tenenboim-Weinblatt, K.: Convergent news? A longitudinal study of similarity and dissimilarity in the domestic and global coverage of the Israeli-Palestinian conflict. J. Commun. **67**(1), 1–25 (2017)
4. Boumans, J.: Outsourcing the news? An empirical assessment of the role of sources and news agencies in the contemporary news landscape (2016)
5. Bro, R.: PARAFAC tutorial and applications. Chemometr. Intell. Lab. Syst. **38**(2), 149–171 (1997)
6. Bro, R., Kiers, H.A.: A new efficient method for determining the number of components in parafac models. J. Chemometr. **17**(5), 274–286 (2003)
7. Clausen, L.: Localizing the global: 'domestication' processes in international news production. Media Cult. Soc. **26**(1), 25–44 (2004)
8. Flew, T., Waisbord, S.: The ongoing significance of national media systems in the context of media globalization. Media Cult. Soc. **37**(4), 620–636 (2015)
9. Galtung, J., Ruge, M.H.: The structure of Foreign news the presentation of the Congo, Cuba and Cyprus crises in four Norwegian newspapers. J. Peace Res. **2**(1), 64–90 (1965)
10. Harcup, T., O'neill, D.: What is news? Galtung and Ruge revisited. J. Stud. **2**(2), 261–280 (2001)

11. Kwak, H., An, J.: A first look at global news coverage of disasters by using the GDELT dataset. In: Aiello, L.M., McFarland, D. (eds.) SocInfo 2014. LNCS, vol. 8851, pp. 300–308. Springer, Cham (2014). doi:10.1007/978-3-319-13734-6_22
12. Kwak, H., An, J.: Multiplex media attention and disregard network among 129 countries. In: Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2017) (2017)
13. Peterson, S.: Foreign news gatekeepers and criteria of newsworthiness. J. Q. **56**(1), 116–25 (1979)
14. Rokach, L., Maimon, O.: Clustering Methods. Springer, Boston (2005)
15. Sreberny-Mohammadi, A.: The "world of the news" study. J. Commun. **34**(1), 121–134 (1984)
16. Wu, H.D.: Systemic determinants of international news coverage: a comparison of 38 countries. J. Commun. **50**(2), 110–130 (2000)
17. Wu, H.D.: Homogeneity around the world? Comparing the systemic determinants of international news flow between developed and developing countries. Int. Commun. Gaz. **65**(1), 9–24 (2003)