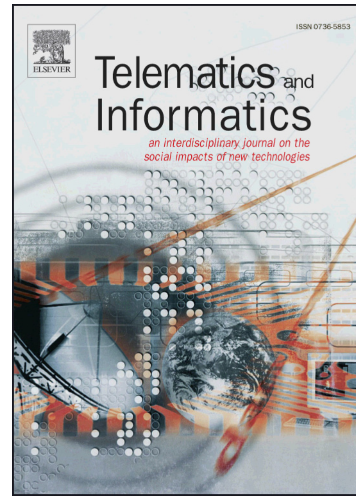
Accepted Manuscript



Social Networking Site Usage and Participation in Protest Activities in 17 Latin–American Countries

Víctor Hugo Masías, Tobias Hecking, Ulrich Hoppe

PII: S0736-5853(18)30360-5
DOI: https://doi.org/10.1016/j.tele.2018.05.010
Reference: TELE 1129

To appear in: *Telematics and Informatics*

Please cite this article as: Masías, V.H., Hecking, T., Hoppe, U., Social Networking Site Usage and Participation in Protest Activities in 17 Latin–American Countries, *Telematics and Informatics* (2018), doi: https://doi.org/10.1016/j.tele.2018.05.010

Social Networking Site Usage and Participation in Protest Activities in 17 Latin–American Countries ☆

Abstract

Recent studies have reported a significant relationship between the use of social media and political engagement. However, there appear to be few comparative studies that explore the association between social networking site use and participation in different types of protest activities for the case of Latin America. The present study employs supervised and unsupervised data analysis techniques to explore this association for 9 different social networking sites and 5 types of protest using disaggregated data on 17 Latin American countries. Multiple correspondence analysis is applied to create proxy measures of the two phenomena, followed by a cluster analysis using these measures to classify individuals into different clusters in each country studied. These clusters indicated that there exists an interplay between the use of these sites and participation in protests. Decision rules were then induced to generate interpretable information on the clusters identified for each country. The results suggest that there is a high degree of heterogeneity in social networking site use and protest participation.

Keywords: Latin America, Multivariate Exploratory Data Analysis, Social Media Usage, Protest activities, Multiple Correspondence Analysis, Clustering, Decision Rule Induction

☆Fully documented templates are available in the elsarticle package on CTAN.

1. Introduction

Latin America has been the setting for a number of civil protest movements over the last decade in which Internet social media have been used as a means of political expression (Guzman-Concha, 2012; Pérez-Liñán and Polga-Hecimovich, 2016; Anselmi, 2017; Valenzuela, 2013; Harlow, 2011). This article explores the relationship between the use of different social networking sites and participation in protest activities in a series of Latin American countries.

Recent comparative studies have shown that there is a positive relationship between social media use and protest participation (Kim and Chen, 2016; Boulianne, 2017; Dong et al., 2017; Gan et al., 2017). Meta-analyses, for example, have found evidence of a relationship between social media use and such participation in both online and offline environments (Skoric et al., 2016a; Boulianne, 2015, 2017; Skoric et al., 2016b). In one of these investigations, based on 133 cross—sectional studies, the author concluded that the effects of using social media on such participation are stronger for political expression and weaker for informational purposes (Boulianne, 2017).

What was not established in these studies, however, was which social media was (were) associated with which type(s) of protest. There thus remains a need to identify the more specific relationships existing between the latter and the many different social networking sites that currently exist. The exploration of individuals' use of such sites, or of the many types of protest they may participate in, is a complex, multidimensional task, however. To be sure, considerable work has already been done on the use by individuals of multiple social media channels. For example, in Hecking et al. (2018, 2017) the authors develop a tool and a strategy for analyzing the dissemination of content through Web pages, Twitter and Wikipedia. Another case is Farahbakhsh et al. (2016), which explores the cross-posting activity of users across three major social networking sites: `Facebook`, `Twitter` and `Google+`. But the research into this phenomenon is complicated by the sheer number of different social media, various problems to do with defining data models, anonymisation and privacy issues, and certain ethical questions (Henderson et al., 2013; Kaslow et al., 2011).

Investigations have also shown that individuals engage in a variety of different types of protest (Mourão et al., 2016; Fourcade et al., 2016; Harlow and Harp, 2012). Traditional protest modes and those that actually bring about changes in social situations still occur in offline environments but it is also true that much protest is now taking place online (Harlow and Harp, 2012). The latter type of activity is frequently hosted by social networking sites. But if, as we have just observed, tracking the use of multiple sites and the participation in different protest modes are complex tasks when tackled separately, the complexities of exploring the relationships between them are even greater.

The present work therefore proposes to take the first step towards filling the gaps in our knowledge on these relationships by investigating the interplay between users of specific social media sites and their participation in specific protest activities in Latin America. Furthermore, our analysis will be conducted at a disaggregated level, that is, country by country. Thus, we will explore the association between participation in 5 types of protest activity and the use of 9 social media sites in 17 different Latin American countries. For each of these countries our objective will be threefold: (i) to explore the patterns of social networking site use and protest participation, (ii) to cluster individuals according to these two concepts, and (iii) to interpret the resulting clusters. To accomplish this we will bring to bear three techniques in turn: (i) multiple correspondence analysis, (ii) agglomerative hierarchical cluster analysis, and (iii) decision rules induction. The first two are data exploration techniques while the third is an automatic supervised learning technique we will apply to derive interpretable information from the clusters obtained using the second technique. This approach will allow us to carry out our analyses in a low-dimensional space.

The remainder of this paper is divided into four main sections. Section 2 reviews existing research; Section 3 describes the proposed methodology, the variables included and the techniques of analysis applied; Section 4 sets out our results; Sections 5 and 6 discuss our findings and some of its limitations; and finally, Section 7 presents our conclusions.

2. Previous research

From a theoretical standpoint, various authors have hypothesized that exposure to the new communications media is positively related to protest participation (Skoric et al., 2016a; Boulianne, 2015, 2017; Skoric et al., 2016b). Social networking sites are a communications media technology that extends both the reach and speed of information dissemination to different audiences (Garrett, 2006). The relatively low cost of social media access today has turned the many existing social media channels into highly efficient alternatives for providing information on political issues, inviting and coordinating the involvement of others in protest activity, and disseminating multimedia content regarding protest activities taking place both online and offline (Kruikemeier and Shehata, 2016; Baym, 2015; Skoric et al., 2011; Steinert-Threlkeld et al., 2015). In this sense, the many social media channels can be used virtually to organize such activities in Latin America.

Various studies have been published on the use of social media in Latin American protest activities in (Pérez-Liñán and Polga-Hecimovich, 2016; Valenzuela, 2013; Harlow, 2011; Mourão et al., 2016) but few have analyzed participation in multiple protest activities and/or the use of multiple social networking sites in these countries. Below we briefly examine a number of previous studies related to the present article. In Harlow and Harp (2012), the authors compared social networking site use and protest participation in the United States with the same phenomena in Latin America, but they also pointed out the need for investigations originating in the region that do not employ an American or Eurocentric perspective on political activism. They reported that in Latin America, 100% of activists surveyed used Facebook to engage in social activism while about 40% used Twitter. In addition, they found that respondents participated to an equal extent in offline protest activities and believed that online activity translated into offline protest. Furthermore, 70% of those interviewed said they felt that social media use played an important part in social movements. The study concluded that "US activists were more likely than those in Latin America to use social networking sites for activism, or to say their activism occurred mostly online" (Harlow and Harp, 2012, p. 10). It did not, however, disaggregate its results by country, comparing U.S. interviewees to Latin American ones as a whole.

Another investigation of patterns of protest in Latin America (Mourão et al. (2016)) divided the individuals surveyed into two types based on whether the protest tactics they employed were moderate or radical. This study is particularly valuable in that the results were disaggregated by country, but it did not gather data on individuals' social networking site use and thus could not examine the relationship between the latter and protest activities. Fourcade et al. (2016) looked into protest activities on various continents including Latin America, gathering data on individuals' membership and participation in associations, demonstrations, strikes, petitions, boycotts and occupations of buildings. The authors showed that the separate countries can be described in terms of the protest activities individuals take part in, but unlike Mourão et al. (2016) they did not consider data either on social networking site use or on online activity and thus did not attempt to present a general vision of the use of different types of social media sites. Similarly, Valenzuela et al. (2016) contains no information on these various sites or their use by the individuals questioned.

In short, previously published research does not seem to have examined the association between (multiple) social networking site use and participation in (multiple) protest activity in Latin America broken down by country (i.e. a disaggregated analysis). The above observations regarding the coverage of these studies are summarized in Table 1.

Table 1: Summary of previous related studies

Author	Level of Analysis	Multiple Social Networking Sites	Multiple Protest Activities	Main Techniques
Harlow and Harp (2012)	Aggregated over countries	Yes	Yes	t test, χ^2 test, correlation analysis, and ANOVAs
Mourão et al. (2016)	Disaggregated by country	No	Yes	Factor analysis, cluster analysis, linear regression
Fourcade et al. (2016)	Aggregated/Disaggregated by country	No	Yes	Principal component analysis, multiple correspondence analysis
Valenzuela et al. (2016)	Disaggregated by country	No	No	Logistic regression

As can be seen, the last column in the table describes the main analytical techniques used in these previously published studies. Harlow and Harp (2012), for example, used a series of exploratory techniques to compare social media user samples in the U.S. and Latin America. Fourcade et al. (2016) applied principal component and multiple correspondence analyses to create proxies for protext participation while Mourão et al. (2016) utilized factor analysis to identify underlying constructs and cluster analysis to segment individuals. Lastly, Valenzuela et al. (2016) attempted to generate knowledge by creating logistic regression models of a binary response variable (participation or non-participation in protests) as a function of various explanatory variables. Thus, we may say that existing methodological approaches are both hetergeneous and exploratory. Furthermore, they share the fact that none of them employ techniques which would facilitate the interpretation of the patterns found in the data.

In light of the foregoing, we propose that by constructing a number of proxy measures for quantifying the use of social media services and participation in protest activities, it should be possible to study the profiles of individuals, group/segment them into clusters, and study these clusters' attribute values.

3. Materials and Methods

This section introduces our methodological approach for exploring the relationship between social networking site use and protest participation in Latin American countries. It also describes the data sample, the variables considered and the techniques applied to analyze the data.

3.1. Methodological approach

Our methodological approach is inspired by the techniques of exploratory multivariate data analysis. This approach was introduced by Tukey (1977) as a way of analyzing data in cases where there is little knowledge or contextual information on the topic under study. The idea is not to make the data fit the researcher's predetermined statistical model or to attempt to predict a predetermined variable, but rather to develop methods for summarizing, describing datasets and creating low-dimensional visualizations.

Thus, we propose to focus on a multivariate exploration of the data and the interpretation of patterns identified in them. The first technique we use is multiple correspondence analysis (Section 3.4.1), which will allow us to create proxies and represent the data in low-dimensional space. The second technique we employ is hierarchical cluster analysis (Section 3.4.2), in order to identify patterns in the data based on the aforementioned proxies. Finally, to interpret the identified patterns we induce decision rules for each cluster (Section 3.4.3). This will allow us to extract key information on the clusters and interpret them.

The methodological approach just outlined is summarized by the flow diagram in Figure 1 illustrating the five steps in the process that was followed for each country in the study. It begins with the selection of a series of variables for measuring social networking site use and protest participation that were obtained from Latinobarómetro, a survey conducted by a non–profit polling NGO based in Santiago, Chile. The second step is the multiple correspondence analysis (MCA), which is used to model the social networking site concept in one dimension and the protest participation concept in another using the survey data on the variables in Step 1. This is followed by the agglomerative hierarchical cluster analysis, using the results of the MCA as input. In the fourth step, the membership labels of the clusters are used as class labels to train a decision rule classifier that then induces a set of decision rules. Finally, the results of these analyses are interpreted. The five steps are described in detail below.

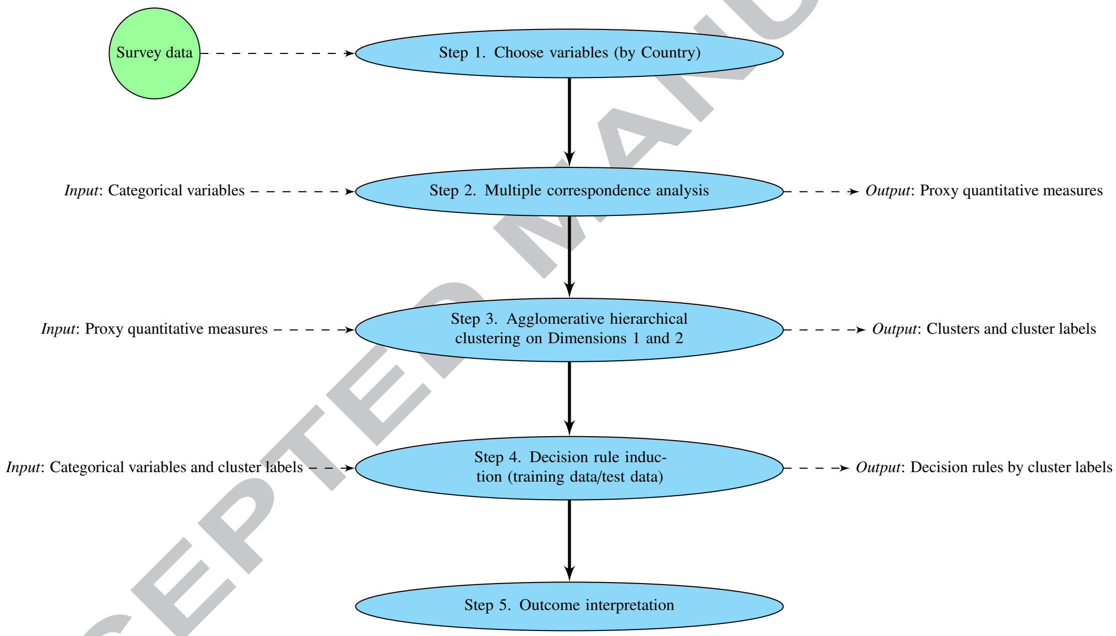


Figure 1: Steps in methodological approach for exploring the relationship between social networking sites use and participation in protest activities disaggregated by country.

3.2. *Description of participants*

The Latinobarómetro survey data related to the period January 15 through February 15, 2015. The total sample size was 19,050 individuals. The survey questionnaire was identical for each of the 17 countries, and to avoid misunderstandings arising from the language and dialect variations across Latin America, interviews in every country were carried out face-to-face by native speakers of the local language variant. The margin of error (sampling error) across the various countries ranged from ± 2.8 to ± 3.1 percentage points. All respondents were guaranteed anonymity. The general characteristics of the sample are set out in Table 2.

Table 2: Sample characteristics by country

Country	Sample size	Sampling error (CI 95%)
Argentina	1,200	± 2.8
Bolivia	1,200	± 2.8
Brazil	1,250	± 2.8
Chile	1,200	± 3.0
Colombia	1,200	± 3.1
Costa Rica	1,000	± 3.1
Dominican Republic	1,000	± 3.1
Ecuador	1,200	± 2.8
El Salvador	1,000	± 3.1
Guatemala	1,000	± 3.1
Honduras	1,000	± 3.1
Nicaragua	1,000	± 3.1
Panama	1,000	± 3.1
Paraguay	1,200	± 2.8
Peru	1,200	± 2.8
Uruguay	1,200	± 2.8
Venezuela	1,200	± 3.0

3.3. Variables

There are no universally accepted definitions in the specialized literature of either protest activities (Opp, 2009; Biggs, 2014), social networking sites (Waheed et al., 2017) or social media usage (Hu and Zhang, 2016). For the purposes of this paper a total of 14 variables were used, 9 of which measured different social networking sites while 5 measured different types of protest[1]. The two classes of variables are described in what follows.

3.3.1. Social networking site use

For these 9 variables, the question put to the interviewees was: *"Do you use any of the following social networking services?"*. They were then read a list of media that included `Facebook`, `Google+`, `Youtube`, `Instagram`, `Twitter`, `LinkedIn`, `Hi5`, `MySpace` and `Sonico`. In each case, the responses were coded as either "`No`" or "`Yes`" as appropriate.

3.3.2. Protest participation

For these 5 variables, the request put to the interviewees was: *"I am going to read out a list of political activities that people can participate in and I would like you to tell me if you have taken part in any of them."* The list of activities included protests in the media (`PM`), protest by signing a petition (`PP`), protest through social media (`PSM`), an officially authorized protest (`PA`) and a non–officially authorized protest (`PNA`). In each case, the responses were coded as either "`No`" or "`Yes`" as appropriate.

3.4. Data analysis

The three main supporting analytical techniques mentioned above are detailed below.

[1]The `missMDA` package developed by Josse et al. (2016) is used for the imputation of missing values in the data sets. The iterative MCA algorithm regularized is used to avoid over-fitting given the number of missing values. Individuals were weighted for each country using the WT variable provided by Latinobarómetro Corporation. All variables were dichotomized to obtain the same number of levels to apply the MCA technique.

3.4.1. Multiple correspondence analysis

Multiple correspondence analysis MCA (MCA) is an exploratory method for summarizing and visualizing multidimensional categorical data. The following description covers the basics of MCA as applied in this paper; for a comprehensive description of the method including a tutorial, the reader is referred to Husson and Josse (2014) and Josse et al. (2012).

Consider a dataset composed of N individuals described by Q categorical variables. Each categorical variable q_k can take J_q categorical values. $J = \sum_{q=1}^{Q} J_q$ thus denotes the total number of different attribute values. A special type of contingency matrix $\mathbf{Z} \in \{0, 1\}^{N \times J}$ is constructed known as the binary indicator matrix, in which the rows are the individuals and the columns are the possible variable attribute values. Every row z_i has Q 1-elements, each one indicating that the corresponding value was one of those chosen for the Q variables. The row entries will thus sum up to Q, the number of categorical variables. Together, the rows in $\mathbf{Z}$ form the cloud of individuals while the columns can be interpreted as the cloud of categories. The fundamental idea behind MCA is to use the indicator matrix as input for dimensionality reduction to fit the cloud of individuals or categories. The diagonal matrix of column margins $\mathbf{C}$ is denoted by D_c. MCA can then be expressed as a weighted principal component analysis (PCA) by the singular value decomposition given in the following equation:

$$\tilde{\mathbf{Z}} = \frac{1}{\sqrt{QN}}(\mathbf{Z} - \tilde{\mathbf{1}}\mathbf{C}^T)D_C^{-\frac{1}{2}} = \mathbf{U}\mathbf{\Lambda}\mathbf{V} \tag{1}$$

An important property of the matrix $\tilde{\mathbf{Z}}$ is that the distance between its rows and columns corresponds to the χ^2 distance $\mathbf{Z}$. The diagonal matrix $\mathbf{\Lambda}$ of the singular values $\tilde{\mathbf{Z}}$ can be used to weight the dimensions (or factors) of $\tilde{\mathbf{Z}}$. Consequently, the factor loadings for the individuals $\mathbf{F}$ as well as the variable values $\mathbf{G}$ can be retrieved using equations 2 and 3, respectively.

$$\mathbf{F} = \frac{1}{\sqrt{N}}\mathbf{U}\Lambda \tag{2}$$

$$\mathbf{G} = D_C^{-\frac{1}{2}}\mathbf{V}\Lambda \tag{3}$$

Here $\mathbf{U}$ and $\mathbf{V}$ are, respectively, the matrices of the left and right singular vectors of the SVD described in Equation 1.

3.4.2. Agglomerative hierarchical clustering

To group the individuals we used an approach developed in Husson et al. (2017), applying it once the raw data were transformed into numerical variables using the MCA as recommended by the same authors. Only two dimensions were included in the analysis so that the models obtained would be interpretable.

The hierarchical clustering analysis itself was performed based on Ward's agglomerative criterion for tree construction (Ward, 1963). The traditional Euclidean distance metric was used given that the dimensions obtained from the MCA can be interpreted with it even if the initial row space distance is a χ^2 distance (Husson et al., 2017). As will be shown below in Subsection 4.1, we chose Dimensions 1 and 2 because they explained the greatest amount of variance in addition to being the most statistically reliable and interpretable.

Various simulations were conducted to find the clusters of individual profiles for each country, restricting the searches to solutions with a number of clusters anywhere from 2 to 10. To validate these clusters we were guided by the dendrogram, the gain of inertia obtained by adding a cluster, and cluster interpretability, as suggested in Renaud-Gentié et al. (2014) and Husson et al. (2017). We also checked a number of statistical characteristics of the clusters. These included the Value–test (V–test), which calculates the contribution of each dimension to cluster formation. This test indicates whether the mean of a cluster is above or below the general mean. The values are considered to be significantly different when V–test >2, *p*–value <0.05 (Lebart et al., 1998; Cornillon et al., 2012).

3.4.3. *Decision rule induction*

The C50 algorithm was applied to induce decision trees (DT) for constructing classifiers that would separate the cluster instances in order to derive interpretable decision rules. This was done using the cluster labels found in the previous step (i.e., the hierarchical cluster analysis).

The C50 (Quinlan, 2003) is a classic algorithm for this purpose and the successor to the C4.5 algorithm (Quinlan, 2014), which in turn superseded the ID3 (Quinlan, 1986). The division rule is a distinctive element in a DT that represents the mechanism by which the instances in a given group form the tree's nodes. It is used to choose the best partition for a predictor variable. The choice criterion is the maximization of a goodness–of–fit measure. In the case of the C50, the splitting criterion is the *information gain*. This measure is based on the probability that an instance belongs to one of the classes. The information gain ratio expresses the ability of a predictor variable to discriminate between instances in any of the *k* possible classes in terms of information gained. For more information on the C50, see the C50 package (Kuhn et al., 2014).

3.5. *Verifying sample size requirements*

Steps were taken to check that the sample size was sufficient for the application of the above-described techniques. For MCA, there are no set rules on sample size but a minimum of 20 cases per variable has been suggested (Franco, 2015). For cluster analysis, one study (Qiu and Joe, 2009) recommends that the sample size be at least 10 times the number of variables, another (Dolnicar et al., 2013) counsels the use of sample sizes 70 times the number of variables and a third (Dolnicar et al., 2016) advocates sample sizes of 10 to 30 times the number of variables to improve cluster recovery. As regards training the classifier, for each country 70% of the cases were used as the training set and 30% as the test set.

4. Results

This section sets out the results of our proposed methodological approach. We begin with those obtained using MCA (Section 4.1), followed by those generated by the agglomerative hierarchical clustering analysis (Section 4.2) and the induced decision rules (Section 4.3).

4.1. *Multiple correspondence analysis results*

In general terms, the exploratory analysis shows it is possible to identify patterns in social networking site use and protest participation. The quantitative results of the `MCA` broken down by country are shown in Table 3. They were obtained using the first two dimensions as suggested in Gifi (1990). As can be seen, the country with the highest percentage of cumulative explained variance was the Dominican Republic while the country with the lowest percentage was Costa Rica. For `Dim 1`, the highest reliability score was registered by Colombia (α = 0.82) and the lowest by Ecuador (α = 0.70). In the case of `Dim 2`, the highest score was attained by the Dominican Republic (α = 0.64) and the lowest by Costa Rica (α = 0.45). The percentage of variance explained and the reliability is adequate if we consider the number of variables included and the attributes of each one.

Table 3: Dimension, explained variance (%), and reliability by country

Country	Dimension	Eigenvalue	% Variance	% Variance (cumul.)	α *
Argentina	`Dim 1`	0.21	20.70	21.00	0.71
	`Dim 2`	0.15	15.00	36.00	0.56
Bolivia	`Dim 1`	0.21	20.90	21.00	0.71
	`Dim 2`	0.16	15.80	37.00	0.59
Brazil	`Dim 1`	0.29	28.70	29.00	0.81
	`Dim 2`	0.15	14.90	44.00	0.56
Chile	`Dim 1`	0.22	21.80	22.00	0.72
	`Dim 2`	0.16	15.70	38.00	0.58
Colombia	`Dim 1`	0.30	29.60	30.00	0.82
	`Dim 2`	0.13	13.30	43.00	0.50
Costa Rica	`Dim 1`	0.22	22.10	22.00	0.73
	`Dim 2`	0.12	12.30	34.00	0.45
Dominican Republic	`Dim 1`	0.28	27.70	28.00	0.80
	`Dim 2`	0.18	17.80	46.00	0.64
Ecuador	`Dim 1`	0.21	20.80	21.00	0.70
	`Dim 2`	0.17	17.00	38.00	0.62
El Salvador	`Dim 1`	0.24	24.00	24.00	0.76
	`Dim 2`	0.13	13.30	37.00	0.50
Guatemala	`Dim 1`	0.23	23.40	23.00	0.75
	`Dim 2`	0.13	12.80	36.00	0.47
Honduras	`Dim 1`	0.23	23.40	23.00	0.75
	`Dim 2`	0.14	13.80	37.00	0.52
Nicaragua	`Dim 1`	0.24	24.20	24.00	0.76
	`Dim 2`	0.13	13.10	37.00	0.49
Panama	`Dim 1`	0.27	26.90	27.00	0.79
	`Dim 2`	0.15	15.50	42.00	0.58
Paraguay	`Dim 1`	0.22	22.10	22.00	0.73
	`Dim 2`	0.14	14.30	36.00	0.54
Peru	`Dim 1`	0.22	21.90	22.00	0.72
	`Dim 2`	0.15	14.60	37.00	0.55
Uruguay	`Dim 1`	0.23	22.90	23.00	0.74
	`Dim 2`	0.15	15.50	38.00	0.58
Venezuela	`Dim 1`	0.23	22.60	23.00	0.74
	`Dim 2`	0.14	13.80	36.00	0.52

**Note*: Cronbach's Alpha (α) for each dimension was computed using the approximation given in Greenacre (2017).

To interpret these results we explore the individuals cloud and the categories cloud corresponding to Dimensions 1 and 2 on cloud graphs for each country in Figure 2 and A1. Each point in magenta is an individual profile and each blue triangle is a category and its value. Positive values for `Dim 1` indicate individuals who use social media while positive values for `Dim 2` indicate those who take part in protests. Notice regarding the individuals clouds that in Quadrant I of the graphs there is a wide variety of profiles whereas in Quadrants II and IV there is less variety and in Quadrant III there is still less.

Although the two dimensions distribute the individual profiles across the four quadrants, suggesting that each one is grouping similar individual profiles and category values, we cannot yet interpret them in the sense of determining which categories distinguish them from each other. For this, we must analyze the categories clouds. As can be seen, the graphs for Argentina, Bolivia, Brazil, Chile, Costa Rica, Ecuador, Honduras, Peru and Venezuela, show that the category values in Quadrant IV (i.e., Facebook = Yes, Google+ = Yes, Youtube = Yes, Instagram = Yes, Twitter = Yes, Linkedin = Yes, Hi5 = Yes, MySpace = Yes, Sonico = Yes) are the opposite of those in Quadrant II (i.e., Facebook = No, Google+ = No, Youtube = No, Instagram = No, Twitter = No, Linkedin = No, Hi5 = No, MySpace = No, Sonico = No, Sonico = No). Also, the categorical values in Quadrant I (i. e. PA = Yes, PP = Yes, PM = Yes, PSM = Yes, PNA = Yes) are the opposite of those in Quadrant III (i.e., PA = No, PP = No, PM = No, PSM = No, PNA = No). As for Colombia, Dominican Republic, El Salvador, Guatemala, Nicaragua, Panama, Paraguay and Uruguay, certain social networking sites that are not very popular are in Quadrant I rather than Quadrant IV, where they tend to be found for the other countries.

Thus, the interpretation of the cloud of individuals and categories suggest that Dim 1 is a proxy for social networking site use and Dim 2 is a proxy for protest participation. In order to verify our interpretation of the results, Appendix graphically reports the relative contributions of the variables to Dimensions 1 and 2 by country (see Figures A2 and A3 respectively).

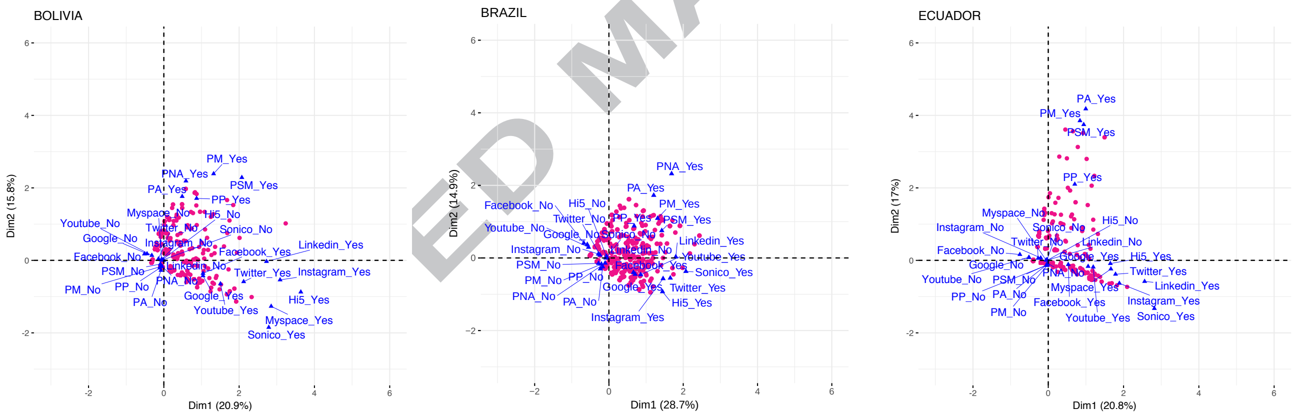


Figure 2: Clouds of individuals (magenta points) and categories (blue triangles) by country (Quadrant I: Dim 1 > 0, Dim 2 > 0; Quadrant II: Dim 1 < 0, Dim 2 > 0; Quadrant III: Dim 1 < 0, Dim 2 < 0; Quadrant IV: Dim 1 > 0, Dim 2 < 0)

4.2. Agglomerative hierarchical clustering results

In this section we present the results of the cluster analysis based on the dimensions generated by `MCA`. Simulations were carried out with each country database to determine the number of clusters K (between 2 and 10 clusters). The characteristics of the clusters found for each country are summarized in Appendix (see Table A1, and Figures A4 and A5).

To demonstrate the usefulness of the results, we examine the clusters obtained for a few of the countries studied. In the case of Bolivia, the solution found contains three clusters. In Cluster 1, both dimensions have values less than the mean (coordinate mean `Dim 1` = –0.21, coordinate mean `Dim 2` = –0.06), a negative and statistically significant V–test value (V–test `Dim 1` = -23.3, V–test `Dim 2` = –8.6) and a significant p–value (p–value `Dim 1` = 1.6e–152, p–value `Dim 2` = 5.0e-172). These results indicate that this cluster groups individuals who neither participate in protests nor use social media sites.

In the three-dimensional Figure 3, which projects the dendrogram for each country onto the two dimensions, Bolivia's Cluster 1 groups homogeneous individuals while Clusters 2 and 3 are more heterogeneous. These features of the three clusters are confirmed by the standard deviation (SD) values, which are much greater for Clusters 2 and 3 than for Cluster 1. The results also show that Cluster 1, grouping those who neither protest nor use social media, has the largest number of individuals (n = 872). Next in size is Cluster 3 (n = 204), representing individuals who use social networking sites but do not participate in protest activities (coordinate mean in category `Dim 1` = 0.73, coordinate mean in category `Dim 2` = -0.31), and the smallest is Cluster 2 (n = 124), containing individuals who clearly do participate in protests but only in offline environments (coordinate mean in category `Dim 1` = 0.28, coordinate mean in category `Dim 2` = 0.94). No group characterizable as individuals who both protest and use social networking sites could be identified.

In the case of Brazil, the individuals can be divided into two groups: those who do not use social media (Cluster 1, n= 772) and those who do (Cluster 2, n= 478). In the latter group, we may expect to find some individuals in Quadrant I who protest on social media (see Figure 3). As for Ecuador, we observe in Figure 3 that Cluster 2 (n=42) contains a small group of individuals in Quadrant I who are participants in different protest activities and use social media (coordinate mean in category `Dim 1` = 0.51, coordinate mean in category `Dim 2` = 1.75).

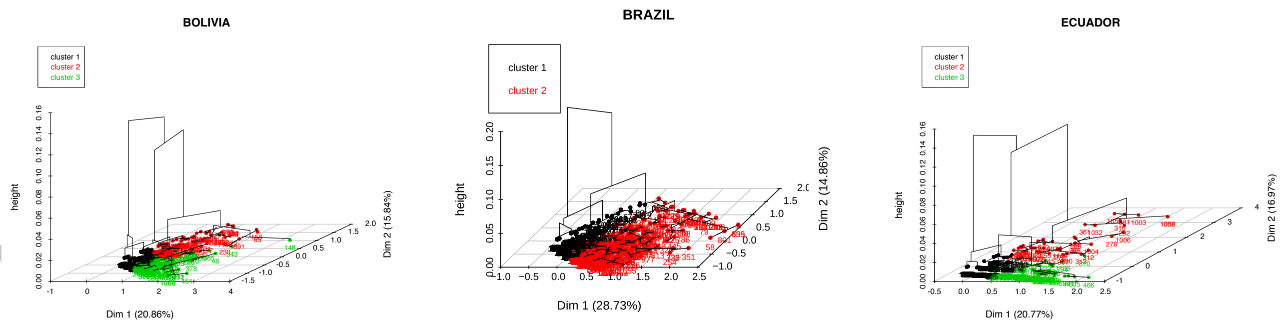


Figure 3: Visualization of agglomerative hierarchical clustering results for Bolivia, Brasil and Ecuador (Quadrant I: `Dim 1` > 0, `Dim 2` > 0; Quadrant II: `Dim 1` < 0, `Dim 2` > 0; Quadrant III: `Dim 1` < 0, `Dim 2` < 0; Quadrant IV: `Dim 1` > 0, `Dim 2` < 0)

It was possible to identify individuals both protesting and using social networking sites for a number of other countries. This was the case with Argentina, Chile, Paraguay, Peru, and Venezuela, which is indicated by their positive coordinate mean values, positive V-test values, and significant p–values, all shown in Table A1. Still, the size of this cluster in these countries tends to be relatively small and heterogeneous. In Venezuela and Ecuador the online activist cluster is virtually undetected. This last observation might seem surprising considering the civil unrest that occurred in Venezuela during 2015, but it must be remembered that in both countries protests were met with government repression and censorship, as has been reported (House, 2015). Two other studies (Ruijgrok, 2016; Boulianne, 2017) arrived at similar results, the latter paper noting that in countries where the press is less than free, individuals tend to use the Internet and/or social media to keep informed but to a lesser degree as a medium of political expression and protest in online environments.

In summary, the number and size of the clusters and the dimensions that statistically characterize them demonstrate that quantitatively, there is considerable variability among the individuals in each country. To obtain a more detailed picture of the patterns of social media use and protest participation from the clusters identified for each country requires requires the application of another set of analytical techniques. This is the objective of the following section.

4.3. Decision rule induction results

In what follows we attempt to show that the information contained in the clusters can be represented by decision rules. For this purpose a classifier is trained using the original database plus the information on which cluster label each individual belongs to. The decision rules are then induced by the `C50` classifier. To validate the decision rule models, the databases were randomly divided into a training sample (70%) and a test sample (30%). The models' performance was measured using Cohen's Kappa coefficient (κ) (Cyr and Francis, 1992).

The results of the validation for each country's decision model are shown in Table 4. In the case of Honduras and Panama, the models performed extremely well while the worst performance was obtained for Nicaragua. The models also did well on the test data, especially for Honduras and Chile, but for Nicaragua once again the results were less impressive. Overall, considering the confidence intervals for κ, the decision rules were very effective in reconstructing the cluster data.

The full set of decision rules for each country, together with the interpretation of them, is provided in Appendix A2. To illustrate the usefulness of the decision rules we compare Bolivia, Brazil and Ecuador, a particularly interesting contrast given that they are geographically close to each other yet culturally different and with different levels of Internet access. Table 5 brings together the decision rules for all of the countries in the study, with each rule consisting of attribute values known as the "rule antecedent" (the "IF" part) and a label called the "rule consequent" (the "THEN" part). In the Bolivian case, Cluster 2 contains individuals who participate in protest activities in offline environments. Rule 3 indicates that individuals who do not use `Youtube` but do take part in both officially authorized and non–officially authorized protests have a high probability of being in Cluster 2 (IF: `Youtube` = `No` $\wedge$ `PA` = `Yes` $\wedge$ `PNA` = `Yes`, THEN: cluster 2). No online activist cluster was identified. In Brazil, on the other hand, users of `Google+` who participate in social media protests (`PSM`) are online activists (IF `Google+` = `Yes` $\wedge$ `PSM` = `Yes`, THEN: Cluster 2). As regards Ecuador, rules 4, 5, 6 and 7 describe the attributes and values that determine whether an individual participates in online and offline protests (i.e., is in Cluster 2).

Table 4: Performance and validation of decision rules by country on training and test data

Dataset	Country	# Instances	Overall Error	κ	SE of κ	95% CI
Training data	Argentina	900	4.3%	0.92	0.01	0.899–0.946
	Bolivia	900	4.9%	0.88	0.01	0.855–0.919
	Brazil	938	2.1%	0.95	0.01	0.935–0.974
	Chile	900	2.6%	0.95	< 0.01	0.938–0.973
	Colombia	900	2.9%	0.94	0.01	0.917–0.963
	Costa Rica	750	3.5%	0.92	0.01	0.897–0.953
	Dominican Republic	750	4.4%	0.91	0.01	0.855–0.942
	Ecuador	900	2.7%	0.93	0.01	0.915–0.963
	El Salvador	750	2.4%	0.91	0.02	0.872–0.952
	Guatemala	750	2.3%	0.91	0.02	0.870–0.953
	Honduras	750	0.7%	0.98	< 0.01	0.962–0.997
	Nicaragua	750	3.1%	0.84	0.03	0.788–0.909
	Panama	750	1.6 %	0.96	0.01	0.938–0.983
	Paraguay	900	3.0%	0.94	0.01	0.919–0.963
	Peru	900	4.8%	0.91	0.01	0.884–0.936
	Uruguay	900	2.4%	0.95	0.01	0.929–0.970
	Venezuela	900	3.0%	0.93	0.01	0.910–0.958
Test data	Argentina	300	4.3%	0.92	0.02	0.878–0.962
	Bolivia	300	6.0%	0.84	0.03	0.782–0.915
	Brazil	312	3.8%	0.91	0.02	0.872–0.963
	Chile	300	2.3%	0.95	0.01	0.929–0.989
	Colombia	300	4.0%	0.91	0.02	0.870–0.963
	Costa Rica	250	4.8%	0.89	0.02	0.841–0.954
	Dominican Republic	250	7.2%	0.86	0.03	0.802–0.922
	Ecuador	300	4.3%	0.88	0.03	0.829–0.946
	El Salvador	250	4.4%	0.84	0.04	0.760–0.935
	Guatemala	250	2.8%	0.87	0.04	0.786–0.966
	Honduras	250	1.2%	0.96	0.02	0.917–1.000
	Nicaragua	250	2.8%	0.81	0.06	0.861–0.948
	Panama	250	4.4%	0.88	0.03	0.815–0.950
	Paraguay	300	3.3%	0.93	0.02	0.889–0.972
	Peru	300	4.7%	0.91	0.02	0.865–0.955
	Uruguay	300	3.7%	0.92	0.02	0.880–0.968
	Venezuela	300	3.3%	0.93	0.02	0.891–0.973

ACCEPTED MANUSCRIPT

Table 5: Induced decision rules for Bolivia, Brazil and Ecuador

Country	Rule	If	Then	Interpretation
Bolivia	Rule 1	`Youtube` = `No` ∧ `Twitter` = `No` ∧ `PP` = `No`	Cluster 1	Neither activist nor social media user
	Rule 2	`Google+` = `No`	Cluster 1	
	Rule 3	`Youtube` = `No` ∧ `PA` = `Yes` ∧ `PNA` = `Yes`	Cluster 2	Offline activist
	Rule 4	`Youtube` = `No` ∧ `PM` = `Yes` ∧ `PP` = `Yes`	Cluster 2	
	Rule 5	`Youtube` = `No` ∧ `PA` = `Yes` ∧ `PP` = `Yes`	Cluster 2	
	Rule 6	`Youtube` = `No` ∧ `PNA` = `Yes`	Cluster 2	
	Rule 7	`Youtube` = `No` ∧ `PP` = `Yes`	Cluster 2	
	Rule 8	`Youtube` = `Yes` ∧ `Twitter` = `Yes`	Cluster 3	Social media user
	Rule 9	`Twitter` = `Yes` ∧ `PP` = `No`	Cluster 3	
	Rule 10	`Youtube` = `Yes`	Cluster 3	
Brazil	Rule 1	`Google+` = `No` ∧ `Instagram` = `No` ∧ `PSM` = `No`	Cluster 1	Non social media activist
	Rule 2	`Facebook` = `No` ∧ `PSM` = `No`	Cluster 1	
	Rule 3	`Youtube` = `No`	Cluster 1	
	Rule 4	`Instagram` = `Yes`	Cluster 2	Social media activist and social media user
	Rule 5	`Google+` = `Yes` ∧ `PSM` = `Yes`	Cluster 2	
	Rule 6	`Twitter` = `Yes`	Cluster 2	
	Rule 7	`Youtube` = `Yes`	Cluster 2	
Ecuador	Rule 1	`Google+` = `No` ∧ `Twitter` = `No` ∧ `Instagram` = `No` ∧ `PNA` = `No` ∧ `PSM` = `No`	Cluster 1	Neither activist nor social media user
	Rule 2	`Facebook` = `No` ∧ `PSM` = `No`	Cluster 1	
	Rule 3	`Youtube` = `No`	Cluster 1	
	Rule 4	`PM` = `Yes` ∧ `PP` = `Yes`	Cluster 2	Online and Offline Activist
	Rule 5	`PA` = `Yes` ∧ `PP` = `Yes`	Cluster 2	
	Rule 6	`PNA` = `Yes`	Cluster 2	
	Rule 7	`Youtube` = `Yes` ∧ `PSM` = `Yes`	Cluster 2	
	Rule 8	`Youtube` = `No` ∧ `Google+` = `Yes` ∧ `Twitter` = `Yes`	Cluster 3	Social media user
	Rule 9	`Youtube` = `Yes`	Cluster 3	

In addition, the rules reveal that the online activist group is heterogeneous given that it does not consist only of individuals with social media accounts who participate in social media protests. On the contrary, using social media is associated with multiple configurations of protest activities. This result is consistent with those reported in Boulianne (2016); Saldana et al. (2015); Towner (2013) and Wolfsfeld et al. (2016), whose authors posit that because social media use exposes people to information about political events, it increases their likelihood of engaging generally in civic activities and political life.

5. Discussion

This paper has explored the association between the use of nine different social networking sites and participation in protest activities for 17 Latin American countries. The results obtained demonstrate that the application of exploratory techniques combined with data mining methods contribute to the identification and interpretation of different groups within the populations of individuals covered in the study.

In addition to this general conclusion, there are a number of more specific aspects that are worthy of comment. First of all, the proposed analysis showed that individuals in each country can be grouped into different clusters. Second, previous research had found that social media site use in protest activities was less common in Latin America than the United States (Harlow and Harp, 2012). The present study has defined more clearly the groups of individuals in Latin American countries that use these sites and take part in protests. Their survey response profiles are generally found in the online activist cluster, a group that is small, heterogeneous and identifiable in certain of the countries explored. The Bolivian case is particularly interesting, however, in that it is the offline activist group, rather than the online activists, that is identifiable. This suggests that in Bolivia the more traditional, offline type of protest activity still predominates.

Third, it was shown that dimensions measuring different concepts can be identified and used to cluster individuals according to those concepts. Thus, employing proxy metrics for social media use and participation in protest activities resulted in an improved approach to the definition of clusters. This we believe constitutes a contribution to current knowledge if contrasted with Mourão et al. (2016), where the authors presented no analysis of clusters by country nor did they include any variables representing specific social networking sites. Here, by contrast, our exploratory analysis used two proxies to represent the use of multiple social networking sites and participation in multiple protest activities in Latin America.

Fourth, the results reported here indicate that clusters of individuals can be interpreted using decision rules. An approach based solely on the quantitative description of such clusters, though suggested in various works on cluster analysis, is not enough to permit a detailed interpretation of the individuals they contain. The approach adopted here demonstrates that the use of automatic learning decision rule methods can deepen the understanding gained through classical statistical analysis. Our methodological approach can also serve as an alternative to explore the general relationship between civic behaviour and the use of social media sites. From the point of view of decision-makers, it is highly useful, since the development of policies that encourage citizen participation and the expression of ideas must be sensitive to the different forms of participation and the social media channels used by a given population.

The final aspect of the present study that deserves mention is the possible effect of political repression and censorship on the number of individuals protesting in social media. This and other repressive measures such as jailing of social media users who expressed political opinions were resorted to in Ecuador and Venezuela during the collection period for the data we used (House, 2015), and in both countries it was observed that the online activist groups were relatively small. These results may indirectly reveal the effects of censorship on social media protests. Individuals in the countries affected might thus confine their use of the Internet and social media to the gathering of information rather than taking part in protest activities (Ruijgrok, 2016; Boulianne, 2017). As noted in Boulianne (2017), "information effects were more likely to be significant in systems without a free press, compared to the United States and other free press systems" (p. 11). Our original expectation was that there would be more rules containing the "PSM=Yes" attribute value (protest through social media). What we found, however, was that this group was made up of individuals who participate in both online and offline protests. This finding is consistent with the idea in Harlow and Harp (2012) that the two broad categories of protest are in fact associated.

6. Limitations

This study has three main limitations. The first one is the relatively low level of variance explained by MCA. Theoretically, increasing this level would improve the reliability of the dimensions. There are two ways this could be done. One is to construct a psychometric instrument designed to measure the concepts under study using known psychometric properties. The other is to utilize more dimensions in the cluster analysis, but in this case the increase in explained variance would be obtained at the cost of greater complexity in interpreting the results. Fortunately, however, the latter alternative is relatively easy to implement and the authors hope to present results using this method in a future publication. For the present article our goal was simply to obtain interpretable results for each of the three stages in the proposed methodology.

The second limitation has to do with our choice of clustering algorithm. Apart from the fact that the agglomerative hierarchical cluster analysis may give different results depending on which distance metric is used, one of its assumptions is that the clusters have no overlap. It would therefore be interesting to apply other clustering methods such as self–organizing maps (Johnsson, 2012) or fuzzy clustering (Miyamoto et al., 2008) that do not make this rather strong assumption.

A third limitation is that the exploration of the individuals was confined to two dimensions. Although, as we have pointed out, the relationship between these two concepts has not been previously studied for Latin America, a more complete analysis of the issue would include additional attributes of the individuals surveyed. The objective we set for this work was limited to two dimensions but the proposed approach could readily be extended to a third or fourth dimension encompassing socio–demographic or socio–economic indicators.

7. Conclusions

This exploratory study attempted to respond to a number of questions regarding the use of social networking sites and protest participation in Latin American countries. The ultimate goal is not merely to explore the relationship between these two phenomena but to undertake the more complex task of constructing causal explanations. We hope that the research reported here has gone some way towards contributing to this purpose.

8. Appendix

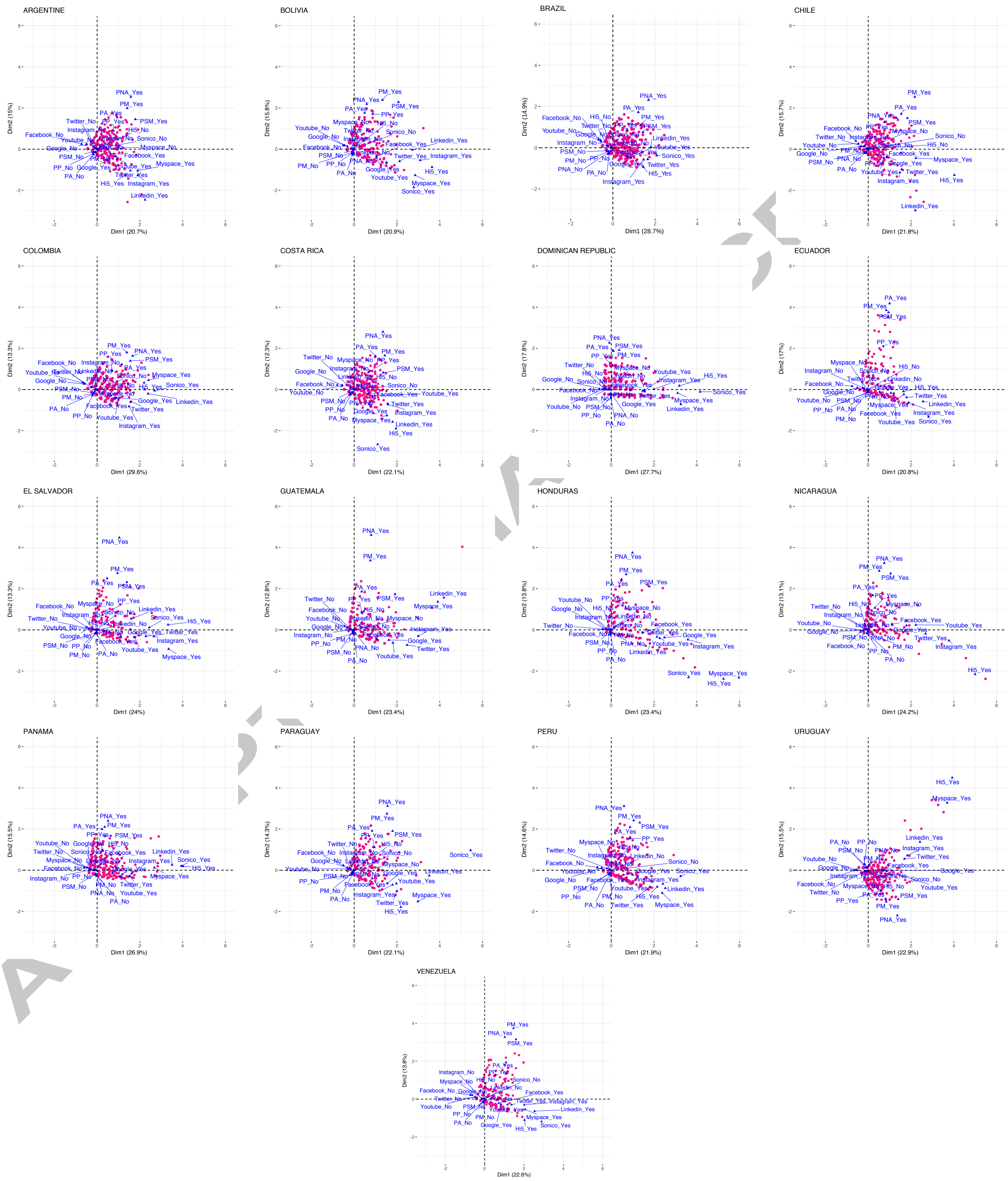


Figure A1: Clouds of individuals (magenta points) and categories (blue triangles) by country (Quadrant I: `Dim 1` > 0, `Dim 2` > 0; Quadrant II: `Dim 1` < 0, `Dim 2` > 0; Quadrant III: `Dim 1` < 0, `Dim 2` < 0; Quadrant IV: `Dim 1` > 0, `Dim 2` < 0)

Figure A2: Relative contribution of categories to Dim 1 by country. The contribution of each category to an axis was considered substantial if it was greater than the expected value under the assumption of a uniform contribution of the variable (R function fvizcontrib; package factoextra).

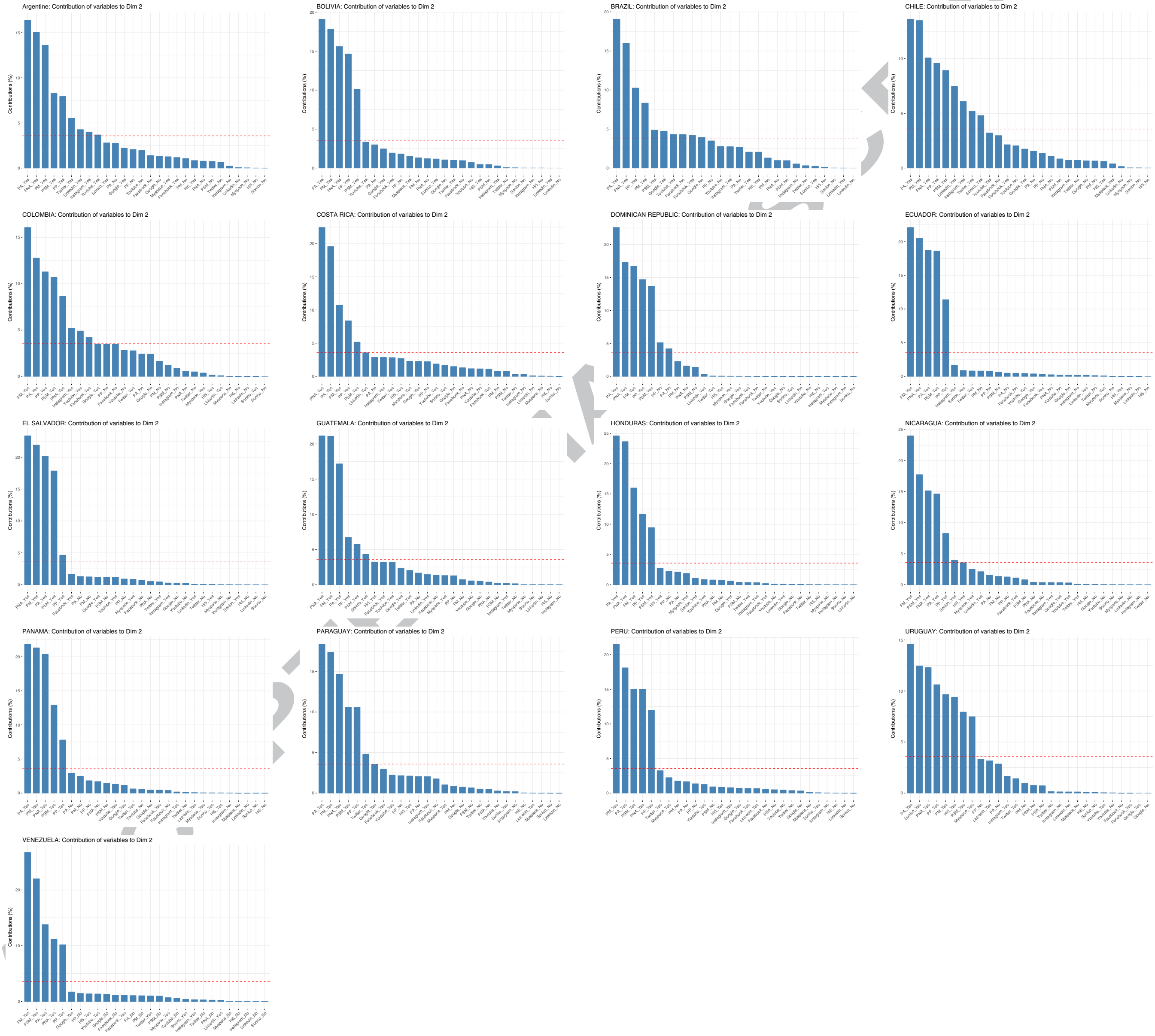


Figure A3: Relative contribution of categories to Dim 2 by country. The contribution of each category to an axis was considered substantial if it was greater than the expected value under the assumption of a uniform contribution of the variable (R function fvizcontrib; package factoextra).

Table A1: Description of clusters by country

Country	*K*	Size	Dim.	V–test	Mean in category	Overall mean	SD in category	Overall SD	*p*–value
Argentina	1	716	`Dim 2`	2.6	0.0055	-0.02	0.15	0.40	9.7e-03
			`Dim 1`	-27.9	-0.2655	0.05	0.19	0.46	1.4e-171
	2	329	`Dim 1`	18	0.41	0.05	0.30	0.46	2.0e-70
			`Dim 2`	-20	-0.39	-0.02	0.27	0.40	1.1e-92
	3	155	`Dim 2`	24	0.67	-0.02	0.4	0.40	4.0e-123
			`Dim 1`	17	0.62	0.05	0.4	0.46	4.3e-64
Bolivia	1	872	`Dim 2`	-8.6	-0.06	2.3e-17	0.14	0.40	7.0e-18
			`Dim 1`	-26.3	-0.21	-2.1e-16	0.17	0.45	1.6e-152
	2	124	`Dim 2`	28.0	0.94	2.3e-17	0.37	0.40	5.0e-172
			`Dim 1`	7.3	0.28	-2.1e-16	0.44	0.45	3.1e-13
	3	204	`Dim 1`	25	0.73	-2.1e-16	0.43	0.45	2.7e-141
			`Dim 2`	-12	-0.31	2.3e-17	0.32	0.40	1.7e-35
Brazil	1	772	`Dim 2`	9.5	0.081	2.9e-16	0.27	0.38	2.5e-21
			`Dim 1`	-28.4	-0.338	7.1e-17	0.21	0.54	3.8e-177
	2	478	`Dim 1`	28.4	0.55	7.1e-17	0.45	0.54	3.8e-177
			`Dim 2`	-9.5	-0.13	2.9e-16	0.49	0.38	2.5e-21
Chile	1	612	`Dim 2`	6.9	0.07	-0.00685	0.11	0.39	5.8e-12
			`Dim 1`	-25.1	-0.34	0.00018	0.10	0.48	1.2e-138
	2	460	`Dim 1`	11	0.21	0.00018	0.37	0.48	1.1e-29
			`Dim 2`	-22	-0.33	-0.00685	0.32	0.39	5.4e-105
	3	128	`Dim 1`	22	0.80	0.00018	0.41	0.48	6.2e-103
			`Dim 2`	22	0.65	-0.00685	0.38	0.39	1.0e-102
Colombia	1	719	`Dim 2`	13	0.092	-0.018	0.27	0.36	4.4e-39
			`Dim 1`	-28	-0.319	0.004	0.21	0.50	2.7e-168
	2	481	`Dim 1`	28	0.50	0.004	0.41	0.50	2.7e-168
			`Dim 2`	-13	-0.19	-0.018	0.42	0.36	4.4e-39
Costa Rica	1	620	`Dim 2`	3.4	0.03	1.4e-16	0.21	0.35	6.4e-04
			`Dim 1`	-26.8	-0.31	-9.5e-17	0.18	0.47	2.7e-158
	2	380	`Dim 1`	26.8	0.509	-9.5e-17	0.33	0.47	2.7e-158
			`Dim 2`	-3.4	-0.048	1.4e-16	0.50	0.35	6.4e-04
Dominican Republic	1	650	`Dim 1`	-18	-0.23	-0.0193	0.17	0.51	2.7e-72
			`Dim 2`	-18	-0.18	0.0021	0.12	0.42	1.8e-75
	2	172	`Dim 2`	26.7	0.79	0.0021	0.36	0.42	5.7e-157
			`Dim 1`	-2.7	-0.11	-0.0193	0.29	0.51	7.5e-03
	3	178	`Dim 1`	25.5	0.90	-0.0193	0.59	0.51	1.9e-143
			`Dim 2`	-3.7	-0.11	0.0021	0.33	0.42	2.3e-04
Ecuador	1	841	`Dim 1`	-30	-0.25	2.9e-17	0.17	0.45	1.9e-191
	2	42	`Dim 2`	28.5	1.75	-2.4e-17	0.84	0.41	6.2e-179

Continued on next page

Table A1 – continued from previous page

Country	*K*	Size	Dim.	V–test	Mean in category	Overall mean	SD in category	Overall SD	*p*–value
			`Dim 1`	7.5	0.51	2.9e-17	0.46	0.45	5.3e-14
	3	316	`Dim 1`	28	0.60	2.9e-17	0.32	0.45	1.4e-166
			`Dim 2`	-10	-0.21	-2.4e-17	0.16	0.41	2.1e-25
El Salvador	1	826	`Dim 2`	4.3	0.025	0.0015	0.33	0.37	1.7e-05
			`Dim 1`	-27.5	-0.196	0.0011	0.16	0.49	2.6e-166
	2	174	`Dim 1`	27.5	0.92	0.0011	0.45	0.49	2.6e-166
			`Dim 2`	-4.3	-0.11	0.0015	0.49	0.37	1.7e-05
Guatemala	1	847	`Dim 2`	7.2	0.037	0.00029	0.31	0.37	8.1e-13
			`Dim 1`	-26.1	-0.160	0.02054	0.17	0.50	3.2e-150
	2	153	`Dim 1`	26.1	0.96	0.02054	0.59	0.50	3.2e-150
			`Dim 2`	-7.2	-0.19	0.00029	0.55	0.37	8.1e-13
Honduras	1	792	`Dim 2`	7.2	0.156	0.067	0.48	0.52	5.3e-13
			`Dim 1`	-24.5	-0.099	0.279	0.16	0.65	1.8e-132
	2	208	`Dim 1`	24.5	0.947	0.279	0.65	0.65	1.8e-132
			`Dim 2`	-7.2	-0.091	0.067	0.55	0.52	5.3e-13
Nicaragua	1	881	`Dim 2`	-7.9	-0.028	0.0069	0.25	0.37	2.6e-15
			`Dim 1`	-24.0	-0.133	0.0097	0.14	0.50	3.3e-127
	2	119	`Dim 1`	24.0	1.02	0.0097	0.85	0.50	3.3e-127
			`Dim 2`	7.9	0.25	0.0069	0.76	0.37	2.6e-15
Panama	1	721	`Dim 2`	6.2	0.048	-9.5e-17	0.38	0.39	5.4e-10
			`Dim 1`	-25.3	-0.257	-4.3e-16	0.16	0.52	1.0e-140
	2	279	`Dim 1`	25.3	0.66	-4.3e-16	0.53	0.52	1.0e-140
			`Dim 2`	-6.2	-0.12	-9.5e-17	0.40	0.39	5.4e-10
Paraguay	1	787	`Dim 1`	-27	-0.24	0.25	0.11	0.61	1.8e-165
	2	134	`Dim 2`	27	0.86	-0.025	0.43	0.46	4.9e-158
			`Dim 1`	12	0.79	0.252	0.61	0.61	5.2e-35
	3	279	`Dim 1`	20	0.70	0.252	0.46	0.61	6.2e-85
			`Dim 2`	-20	-0.37	-0.025	0.26	0.46	2.1e-89
Peru	1	708	`Dim 2`	-2.1	-0.02	2.4e-16	0.13	0.38	3.2e-02
			`Dim 1`	-27.4	-0.31	1.7e-16	0.11	0.47	9.5e-166
	2	385	`Dim 1`	22	0.44	1.7e-16	0.40	0.47	4.5e-110
			`Dim 2`	-15	-0.24	2.4e-16	0.22	0.38	5.2e-50
	3	107	`Dim 2`	28	0.99	2.4e-16	0.42	0.38	3.2e-173
			`Dim 1`	11	0.47	1.7e-16	0.50	0.47	2.2e-27
Uruguay	1	694	`Dim 2`	7.5	0.073	4.6e-17	0.14	0.39	5.6e-14
			`Dim 1`	-26.7	-0.314	5.1e-17	0.15	0.48	9.5e-157
	2	506	`Dim 1`	26.7	0.43	5.1e-17	0.43	0.48	9.5e-157
			`Dim 2`	-7.5	-0.10	4.6e-17	0.57	0.39	5.6e-14
Venezuela	1	795	`Dim 2`	2.1	0.016	1.3e-16	0.16	0.37	3.6e-02

Continued on next page

Table A1 – continued from previous page

Country	*K*	Size	Dim.	V–test	Mean in category	Overall mean	SD in category	Overall SD	*p*–value
			`Dim 1`	-28.9	-0.283	1.7e-16	0.19	0.47	1.4e-183
	2	354	`Dim 1`	25	0.53	1.7e-16	0.31	0.47	8.6e-137
			`Dim 2`	-14	-0.23	1.3e-16	0.22	0.37	1.7e-45
	3	51	`Dim 2`	27	1.38	1.3e-16	0.45	0.37	9.7e-162
			`Dim 1`	11	0.75	1.7e-16	0.56	0.47	2.3e-30

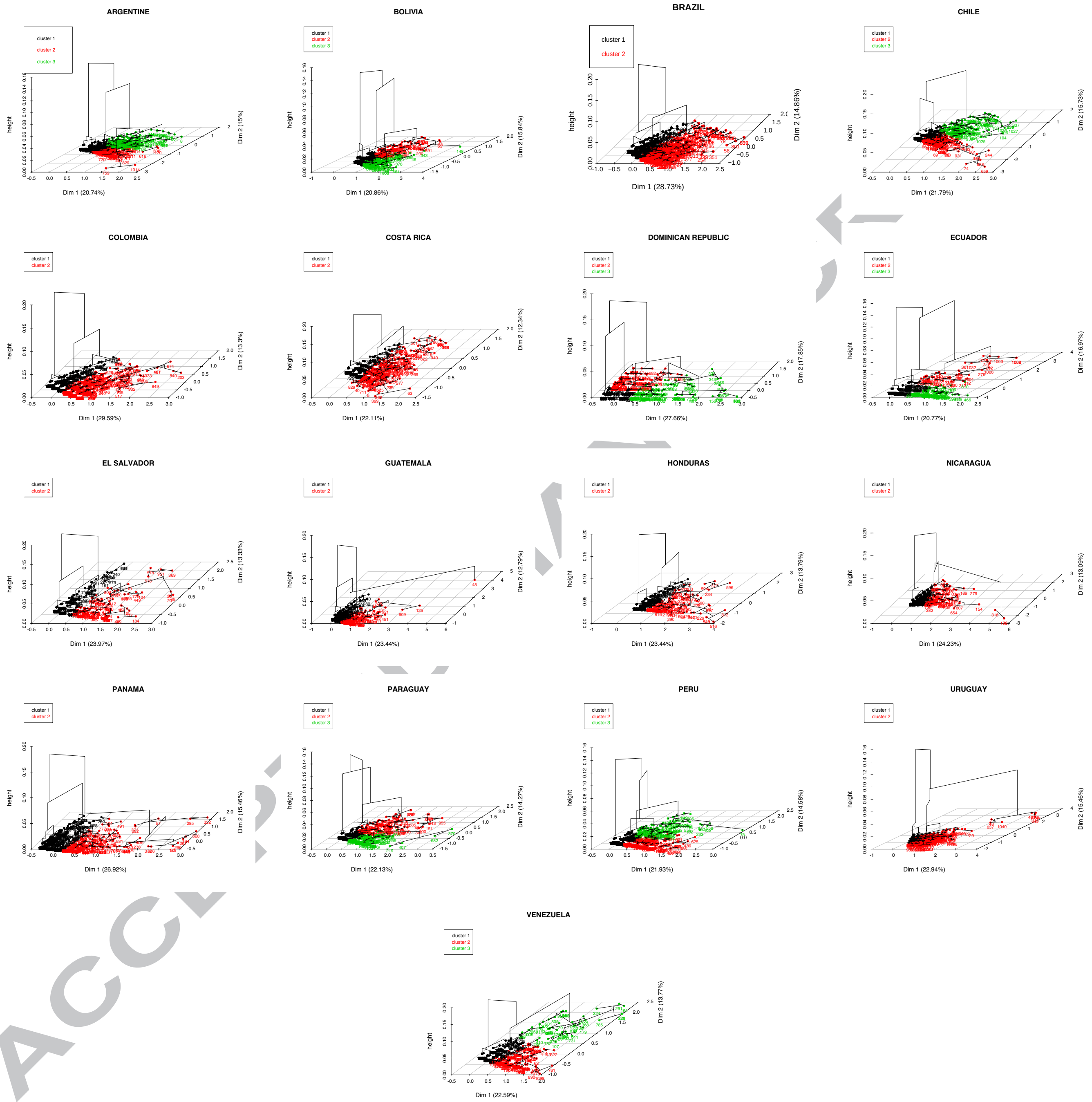


Figure A4: 3D visualization of agglomerative hierarchical clustering (Dim 1, Dim 2, and Dendogram) results by country (Quadrant I: `Dim 1` > 0, `Dim 2` > 0; Quadrant II: `Dim 1` < 0, `Dim 2` > 0; Quadrant III: `Dim 1` < 0, `Dim 2` < 0; Quadrant IV: `Dim 1` > 0, `Dim 2` < 0)

Table A2: Interpretation of decision rules by country

Country	Rule	Lift	If	Then	Conf.	Interpretation
Argentina	Rule 1	(307, lift 1.7)	Facebook = No ∧ Twitter = No ∧ PNA = No ∧ PM = No ∧ PSM = No	Cluster 1	[0.997]	Neither activist nor social media user
	Rule 2	(396/4, lift 1.7)	Youtube = No ∧ Twitter = No ∧ PNA = No ∧ PSM = No ∧ PP = No	Cluster 1	[0.987]	
	Rule 3	(415/6, lift 1.7)	Youtube = No ∧ Twitter = No ∧ PA = No ∧ PNA = No ∧ PSM = No	Cluster 1	[0.983]	
	Rule 4	(542/92, lift 1.4)	Google+ = No	Cluster 1	[0.829]	
	Rule 5	(99/12, lift 3.2)	Twitter = Yes	Cluster 2	[0.871]	Social media user
	Rule 6	(348/126, lift 2.3)	Youtube = Yes	Cluster 2	[0.637]	
	Rule 7	(45, lift 7.6)	Facebook = Yes ∧ Twitter = No ∧ PA = Yes ∧ PP = Yes	Cluster 3	[0.979]	Online and Offline Activist
	Rule 8	(43, lift 7.6)	PNA = Yes	Cluster 3	[0.978]	
	Rule 9	(31, lift 7.5)	PA = Yes ∧ PM = Yes	Cluster 3	[0.970]	
	Rule 10	(25/3, lift 6.6)	Facebook = Yes ∧ Youtube = Yes ∧ PM = Yes	Cluster 3	[0.852]	
	Rule 11	(56/13, lift 5.9)	Twitter = No ∧ PSM = Yes	Cluster 3	[0.759]	
Bolivia	Rule 1	(644/39, lift 1.3)	Youtube = No ∧ Twitter = No ∧ PP = No	Cluster 1	[0.938]	Neither activist nor social media user
	Rule 2	(626/86, lift 1.2)	Google+ = No	Cluster 1	[0.861]	
	Rule 3	(36, lift 9.4)	Youtube = No ∧ PA = Yes ∧ PNA = Yes	Cluster 2	[0.974]	Offline activist
	Rule 4	(26, lift 9.3)	Youtube = No ∧ PM = Yes ∧ PP = Yes	Cluster 2	[0.964]	
	Rule 5	(47/1, lift 9.3)	Youtube = No ∧ PA = Yes ∧ PP = Yes	Cluster 2	[0.959]	
	Rule 6	(53/7, lift 8.3)	Youtube = No ∧ PNA = Yes	Cluster 2	[0.855]	
	Rule 7	(79/22, lift 6.9)	Youtube = No ∧ PP = Yes	Cluster 2	[0.716]	
	Rule 8	(41/1, lift 5.4)	Youtube = Yes ∧ Twitter = Yes	Cluster 3	[0.953]	Social Media User
	Rule 9	(51/2, lift 5.3)	Twitter = Yes ∧ PP = No	Cluster 3	[0.943]	
	Rule 10	(160/31, lift 4.5)	Youtube = Yes	Cluster 3	[0.802]	
Brazil	Rule 1	(424/5, lift 1.6)	Google+ = No ∧ Instagram = No ∧ PSM = No	Cluster 1	[0.986]	Non social media activist
	Rule 2	(412/11, lift 1.6)	Facebook = No ∧ PSM = No	Cluster 1	[0.971]	
	Rule 3	(569/43, lift 1.5)	Youtube = No	Cluster 1	[0.923]	
	Rule 4	(152/2, lift 2.6)	Instagram = Yes	Cluster 2	[0.981]	Social media activist and social media user
	Rule 5	(115/2, lift 2.6)	Google+ = Yes ∧ PSM = Yes	Cluster 2	[0.974]	
	Rule 6	(122/3, lift 2.5)	Twitter = Yes	Cluster 2	[0.968]	
	Rule 7	(369/54, lift 2.2)	Youtube = Yes	Cluster 2	[0.852]	
Chile	Rule 1	(311/1, lift 2.0)	Facebook = No ∧ Google+ = No ∧ PA = No ∧ PM = No	Cluster 1	[0.994]	Neither activist nor social media user
	Rule 2	(536/93, lift 1.6)	Youtube = No	Cluster 1	[0.825]	
	Rule 3	(156/3, lift 2.6)	Youtube = Yes ∧ Google+ = Yes ∧ PA = No ∧ PM = No	Cluster 2	[0.975]	Social media user
	Rule 4	(255/6, lift 2.6)	Facebook = Yes ∧ Youtube = Yes ∧ PA = No ∧ PM = No	Cluster 2	[0.973]	
	Rule 5	(26, lift 2.5)	Facebook = Yes ∧ Youtube = No ∧ Google+ = Yes ∧ PA = No	Cluster 2	[0.964]	
	Rule 6	(38/1, lift 2.5)	Twitter = No ∧ Instagram = Yes ∧ PA = No	Cluster 2	[0.950]	
	Rule 7	(278/20, lift 2.4)	Youtube = Yes ∧ PNA = No ∧ PSM = No	Cluster 2	[0.925]	
	Rule 8	(111/18, lift 2.2)	Twitter = Yes	Cluster 2	[0.832]	
	Rule 9	(47, lift 8.9)	PA = Yes ∧ PNA = Yes	Cluster 3	[0.980]	Online and Offline Activist
	Rule 10	(44, lift 8.9)	PA = Yes ∧ PSM = Yes	Cluster 3	[0.978]	
	Rule 11	(29, lift 8.8)	Youtube = Yes ∧ PA = Yes ∧ PP = Yes	Cluster 3	[0.968]	
	Rule 12	(23, lift 8.7)	Youtube = Yes ∧ PM = Yes	Cluster 3	[0.960]	
	Rule 13	(65/9, lift 7.7)	Youtube = Yes ∧ PA = Yes	Cluster 3	[0.851]	
Colombia	Rule 1	(416/6, lift 1.6)	Google+ = No ∧ Twitter = No ∧ PA = No	Cluster 1	[0.983]	Neither activist nor social media user
	Rule 2	(536/29, lift 1.6)	Youtube = No ∧ Instagram = No	Cluster 1	[0.944]	
	Rule 3	(116, lift 2.5)	Instagram = Yes	Cluster 2	[0.992]	Social media activist and social media user
	Rule 4	(124/3, lift 2.4)	Twitter = Yes	Cluster 2	[0.968]	
	Rule 5	(63/2, lift 2.4)	Google+ = Yes ∧ PSM = Yes	Cluster 2	[0.954]	
	Rule 6	(357/30, lift 2.3)	Youtube = Yes	Cluster 2	[0.914]	
Costa Rica	Rule 1	(349/2, lift 1.6)	Youtube = No ∧ Google+ = No ∧ Twitter = No ∧ PSM = No	Cluster 1	[0.991]	Neither activist nor social media user
	Rule 2	(371/3, lift 1.6)	Youtube = No ∧ Twitter = No ∧ PA = No ∧ PSM = No	Cluster 1	[0.989]	
	Rule 3	(348/7, lift 1.6)	Facebook = No	Cluster 1	[0.977]	
	Rule 4	(378/25, lift 1.5)	Google+ = No ∧ PP = No	Cluster 1	[0.932]	
	Rule 5	(56, lift 2.6)	Google+ = Yes ∧ PSM = Yes	Cluster 2	[0.983]	Online and Offline activist and social media user
	Rule 6	(46, lift 2.6)	Google+ = Yes ∧ PA = Yes	Cluster 2	[0.979]	
	Rule 7	(75/5, lift 2.5)	Twitter = Yes	Cluster 2	[0.922]	
	Rule 8	(283/42, lift 2.3)	Youtube = Yes	Cluster 2	[0.849]	
Dominican Republic	Rule 1	(412/10, lift 1.5)	Youtube = No ∧ PA = No ∧ PSM = No ∧ PP = No	Cluster 1	[0.973]	Neither activist nor social media user
	Rule 2	(604/124, lift 1.2)	Youtube = No	Cluster 1	[0.794]	
	Rule 3	(50, lift 5.7)	Youtube = No ∧ PA = Yes ∧ PNA = Yes	Cluster 2	[0.981]	Online and Offline Activist
	Rule 4	(80/1, lift 5.7)	Youtube = No ∧ PA = Yes ∧ PP = Yes	Cluster 2	[0.976]	
	Rule 5	(33, lift 5.7)	Youtube = No ∧ PM = Yes ∧ PSM = Yes	Cluster 2	[0.971]	
	Rule 6	(74/13, lift 4.8)	Twitter = No ∧ PM = Yes	Cluster 2	[0.816]	
	Rule 7	(74/2, lift 5.4)	Youtube = Yes ∧ Twitter = Yes	Cluster 3	[0.961]	Social media user
	Rule 8	(128/14, lift 5.0)	Youtube = Yes ∧ PM = No	Cluster 3	[0.885]	
Ecuador	Rule 1	(507/7, lift 1.4)	Google+ = No ∧ Twitter = No ∧ Instagram = No ∧ PNA = No ∧ PSM = No	Cluster 1	[0.984]	Neither activist nor social media user
	Rule 2	(313/5, lift 1.4)	Facebook = No ∧ PSM = No	Cluster 1	[0.981]	
	Rule 3	(618/45, lift 1.3)	Youtube = No	Cluster 1	[0.926]	
	Rule 4	(15, lift 22.3)	PM = Yes ∧ PP = Yes	Cluster 2	[0.941]	Online and Offline Activist
	Rule 5	(14, lift 22.2)	PA = Yes ∧ PP = Yes	Cluster 2	[0.938]	
	Rule 6	(11, lift 21.9)	PNA = Yes	Cluster 2	[0.923]	

Continued on next page

Table A2 – continued from previous page

Country	Rule	Lift	If	Then	Conf.	Interterpretation
	Rule 7	(12/2, lift 18.6)	Youtube = Yes ∧ PSM = Yes	Cluster 2	[0.786]	
	Rule 8	(15, lift 3.5)	Youtube = No ∧ Google+ = Yes ∧ Twitter = Yes	Cluster 3	[0.941]	Social media user
	Rule 9	(282/65, lift 2.9)	Youtube = Yes	Cluster 3	[0.768]	
El Salvador	Rule 1	(605/7, lift 1.2)	Google+ = No ∧ Twitter = No	Cluster 1	[0.987]	Non social media user
	Rule 2	(615/10, lift 1.2)	Youtube = No ∧ Instagram = No	Cluster 1	[0.982]	
	Rule 3	(29, lift 5.8)	Instagram = Yes	Cluster 2	[0.968]	Social media user
	Rule 4	(130/19, lift 5.1)	Youtube = Yes	Cluster 2	[0.848]	
Guatemala	Rule 1	(598/8, lift 1.2)	Google+ = No ∧ Twitter = No	Cluster 1	[0.985]	Non social media user
	Rule 2	(623/9, lift 1.2)	Youtube = No ∧ Twitter = No	Cluster 1	[0.984]	
	Rule 3	(90/2, lift 6.1)	Youtube = Yes ∧ Google+ = Yes	Cluster 2	[0.967]	Social media user
	Rule 4	(51/1, lift 6.1)	Twitter = Yes	Cluster 2	[0.962]	
Honduras	Rule 1	(594/4, lift 1.3)	Youtube = No ∧ Google+ = No ∧ Twitter = No	Cluster 1	[0.992]	Non social media user
	Rule 2	(86, lift 4.7)	Google+ = Yes	Cluster 2	[0.989]	Social media user
	Rule 3	(58, lift 4.6)	Twitter = Yes	Cluster 2	[0.983]	
	Rule 4	(116/1, lift 4.6)	Youtube = Yes	Cluster 2	[0.983]	
Nicaragua	Rule 1	(656/14, lift 1.1)	Youtube = No ∧ Twitter = No ∧ Instagram = No	Cluster 1	[0.977]	Non social media user
	Rule 2	(649/19, lift 1.1)	Google+ = No	Cluster 1	[0.969]	
	Rule 3	(21, lift 7.6)	Instagram = Yes	Cluster 2	[0.957]	Social media user
	Rule 4	(32/3, lift 7.0)	Twitter = Yes	Cluster 2	[0.882]	
	Rule 5	(78/10, lift 6.8)	Youtube = Yes	Cluster 2	[0.863]	
Panama	Rule 1	(460/1, lift 1.4)	Youtube = No ∧ Google+ = No	Cluster 1	[0.996]	Non social media user
	Rule 2	(478/2, lift 1.4)	Google+ = No ∧ Twitter = No ∧ Instagram = No	Cluster 1	[0.994]	
	Rule 3	(495/4, lift 1.4)	Youtube = No ∧ Instagram = No	Cluster 1	[0.990]	
	Rule 4	(118, lift 3.5)	Youtube = Yes ∧ Twitter = Yes	Cluster 2	[0.992]	Social media user
	Rule 5	(90, lift 3.5)	Youtube = Yes ∧ Instagram = Yes	Cluster 2	[0.989]	
	Rule 6	(79, lift 3.5)	Google+ = Yes ∧ Instagram = Yes	Cluster 2	[0.988]	
	Rule 7	(167/5, lift 3.4)	Youtube = Yes ∧ Google+ = Yes	Cluster 2	[0.964]	
Paraguay	Rule 1	(732/153, lift 1.2)	Google+ = No	Cluster 1	[0.790]	Non social media user
	Rule 2	(49, lift 8.5)	Google+ = No ∧ PA = Yes ∧ PP = Yes	Cluster 2	[0.980]	Offline activist
	Rule 3	(25, lift 8.3)	PA = Yes ∧ PNA = Yes	Cluster 2	[0.963]	
	Rule 4	(51/4, lift 7.8)	Instagram = No ∧ PM = Yes	Cluster 2	[0.906]	
	Rule 5	(58/7, lift 7.5)	PM = Yes	Cluster 2	[0.867]	
	Rule 6	(70/1, lift 4.1)	Twitter = Yes ∧ PA = No ∧ PM = No	Cluster 3	[0.972]	Offline activist and social media user
	Rule 7	(146/7, lift 4.0)	Google+ = Yes ∧ PNA = No ∧ PM = No	Cluster 3	[0.946]	
	Rule 8	(38/2, lift 3.9)	Google+ = No ∧ Instagram = Yes	Cluster 3	[0.925]	
	Rule 9	(118/14, lift 3.7)	Youtube = Yes	Cluster 3	[0.875]	
Peru	Rule 1	(491/20, lift 1.6)	Youtube = No ∧ Google+ = No ∧ Twitter = No ∧ PA = No	Cluster 1	[0.957]	Neither activist nor social media user
	Rule 2	(474/32, lift 1.6)	Facebook = No ∧ Youtube = No	Cluster 1	[0.931]	
	Rule 3	(199/7, lift 3.0)	Youtube = Yes ∧ PA = No ∧ PSM = No	Cluster 2	[0.960]	Social media user
	Rule 4	(193/8, lift 3.0)	Youtube = Yes ∧ PM = No ∧ PP = No	Cluster 2	[0.954]	
	Rule 5	(15, lift 3.0)	Youtube = No ∧ Twitter = Yes ∧ PA = No	Cluster 2	[0.941]	
	Rule 6	(200/12, lift 2.9)	Facebook = Yes ∧ Google+ = Yes ∧ PP = No	Cluster 2	[0.936]	
	Rule 7	(28, lift 10.2)	PM = Yes ∧ PSM = Yes	Cluster 3	[0.967]	Offline and online activist
	Rule 8	(22, lift 10.1)	PSM = Yes ∧ PP = Yes	Cluster 3	[0.958]	
	Rule 9	(37/3, lift 9.5)	PA = Yes ∧ PP = Yes	Cluster 3	[0.897]	
	Rule 10	(6, lift 9.3)	Facebook = Yes ∧ Youtube = No ∧ Google+ = Yes ∧ PP = Yes	Cluster 3	[0.875]	
	Rule 11	(54/11, lift 8.3)	PSM = Yes	Cluster 3	[0.786]	
	Rule 12	(38/9, lift 7.9)	Google+ = No ∧ PA = Yes	Cluster 3	[0.750]	
Uruguay	Rule 1	(434/6, lift 1.7)	Youtube = No ∧ Twitter = No ∧ PA = No ∧ PNA = No ∧ PSM = No	Cluster 1	[0.984]	Neither activist nor social media user
	Rule 2	(397/10, lift 1.7)	Google+ = No ∧ Twitter = No ∧ Instagram = No ∧ PP = No	Cluster 1	[0.972]	
	Rule 3	(347/11, lift 1.7)	Facebook = No ∧ PP = No	Cluster 1	[0.966]	
	Rule 4	(466/17, lift 1.7)	Youtube = No ∧ Twitter = No ∧ PNA = No ∧ PM = No	Cluster 1	[0.962]	
	Rule 5	(249, lift 2.4)	Facebook = Yes ∧ Youtube = Yes ∧ Google+ = Yes	Cluster 2	[0.996]	Online and offline activist and social media user
	Rule 6	(113, lift 2.3)	Youtube = Yes ∧ PP = Yes	Cluster 2	[0.991]	
	Rule 7	(53, lift 2.3)	Youtube = Yes ∧ Instagram = Yes	Cluster 2	[0.982]	
	Rule 8	(49, lift 2.3)	PM = Yes ∧ PSM = Yes	Cluster 2	[0.980]	
	Rule 9	(77/1, lift 2.3)	Google+ = Yes ∧ PA = Yes	Cluster 2	[0.975]	
	Rule 10	(88/2, lift 2.3)	Twitter = Yes	Cluster 2	[0.967]	
	Rule 11	(50/1, lift 2.3)	PNA = Yes	Cluster 2	[0.962]	
	Rule 12	(83/15, lift 1.9)	PM = Yes	Cluster 2	[0.812]	
Venezuela	Rule 1	(560/10, lift 1.5)	Youtube = No ∧ Twitter = No ∧ PM = No	Cluster 1	[0.980]	Neither activist nor social media user
	Rule 2	(477/9, lift 1.5)	Google+ = No ∧ Twitter = No ∧ PM = No	Cluster 1	[0.979]	
	Rule 3	(467/10, lift 1.5)	Youtube = No ∧ Google+ = No ∧ PM = No	Cluster 1	[0.977]	
	Rule 4	(444/13, lift 1.4)	Facebook = No ∧ PM = No ∧ PSM = No	Cluster 1	[0.969]	
	Rule 5	(128/5, lift 3.3)	Youtube = Yes ∧ Twitter = Yes ∧ PM = No	Cluster 2	[0.954]	Offline activist and social media user
	Rule 6	(164/8, lift 3.2)	Google+ = Yes ∧ Twitter = Yes ∧ PM = No	Cluster 2	[0.946]	
	Rule 7	(209/14, lift 3.2)	Youtube = Yes ∧ Google+ = Yes ∧ PM = No	Cluster 2	[0.929]	
	Rule 8	(29/2, lift 23.9)	PM = Yes	Cluster 3	[0.903]	

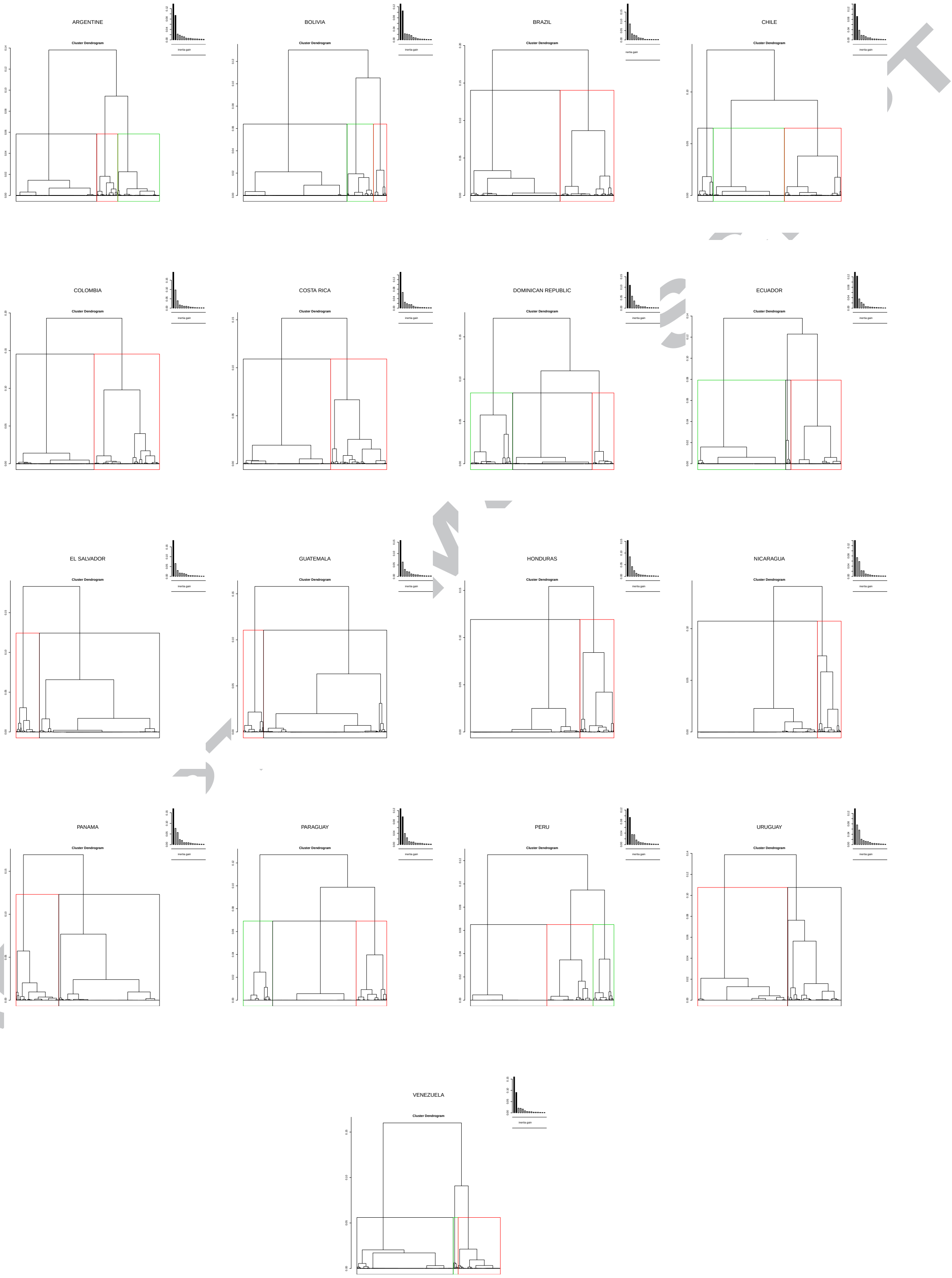


Figure A5: Visualization of agglomerative hierarchical cluster tree and inertia gain

References

Anselmi M. Post-populism in Latin America: On Venezuela after Chávez. Chinese Political Science Review 2017;2(3):410–26. doi:10.1007/s41111-017-0066-y.

Baym N. Personal Connections in the Digital Age. Digital Media and Society. Wiley, 2015. URL: https://books.google.de/books?id=4_1RCgAAQBAJ.

Biggs M. Has protest increased since the 1970s? How a survey question can construct a spurious trend. The British Journal of Sociology 2014;66(1):141–62. doi:10.1111/1468-4446.12099.

Boulianne S. Social media use and participation: A meta-analysis of current research. Information, Communication & Society 2015;18(5):524–38. doi:10.1080/1369118x.2015.1008542.

Boulianne S. Online news, civic awareness, and engagement in civic and political life. New Media & Society 2016;18(9):1840–56. doi:10.1177/1461444815616222.

Boulianne S. Revolution in the making? Social media effects across the globe. Information, Communication & Society 2017;:1–16doi:10.1080/1369118x.2017.1353641.

Cornillon PA, Guyader A, Husson F, Jégou N, Josse J, Kloareg M, Matzner-Lober E, Rouvière L. R for Statistics. CRC Press, 2012. URL: https://books.google.de/books?id=41rNBQAAQBAJ.

Cyr L, Francis K. Measures of clinical agreement for nominal and categorical data: The Kappa coefficient. Computers in Biology and Medicine 1992;22(4):239–46. doi:10.1016/0010-4825(92)90063-s.

Dolnicar S, Grün B, Leisch F. Increasing sample size compensates for data problems in segmentation studies. Journal of Business Research 2016;69(2):992–9. doi:10.1016/j.jbusres.2015.09.004.

Dolnicar S, Grün B, Leisch F, Schmidt K. Required sample sizes for data-driven market segmentation analyses in tourism. Journal of Travel Research 2013;53(3):296–306. doi:10.1177/0047287513496475.

Dong T, Liang C, He X. Social media and internet public events. Telematics and Informatics 2017;34(3):726–39. doi:10.1016/j.tele.2016.05.024.

Farahbakhsh R, Cuevas A, Crespi N. Characterization of cross-posting activity for professional users across facebook, twitter and google+. Springer; volume 6; 2016. p. 33.

Fourcade M, Lande B, Schofer E. Political space and the space of polities: Doing politics across nations. Poetics 2016;55:1–18. doi:10.1016/j.poetic.2015.12.002.

Franco GD. Multiple correspondence analysis: One only or several techniques? Quality & Quantity 2015;50(3):1299–315. doi:10.1007/s11135-015-0206-0.

Gan C, Lee FL, Li Y. Social media use, political affect, and participation among university students in urban China. Telematics and Informatics 2017;34(7):936–47. doi:10.1016/j.tele.2017.04.002.

Garrett RK. Protest in an information society: a review of literature on social movements and new ICTs. Information, Communication & Society 2006;9(2):202–24. doi:10.1080/13691180600630773.

Gifi A. Nonlinear multivariate analysis. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Wiley, 1990. URL: https://books.google.de/books?id=XkamAAAAIAAJ.

Greenacre M. Correspondence Analysis in Practice, Third Edition. Chapman & Hall/CRC Interdisciplinary Statistics. CRC Press, 2017. URL: https://books.google.de/books?id=lCQNDgAAQBAJ.

Guzman-Concha C. The students' rebellion in Chile: Occupy protest or classic social movement? Social movement studies 2012;11(3-4):408–15. doi:https://doi.org/10.1080/14742837.2012.710748.

Harlow S. Social media and social movements: Facebook and an online Guatemalan justice movement that moved offline. New Media & Society 2011;14(2):225–43. doi:10.1177/1461444811410408.

Harlow S, Harp D. Collective action on the web: A cross-cultural study of social networking sites and online and offline activism in the United States and Latin America. Information, Communication & Society 2012;15(2):196–216. doi:10.1080/1369118x.2011.591411.

Hecking T, Steinert L, Leßmann S, Masias VH, Hoppe HU. Identifying accelerators of information diffusion across social media channels. 4th European Network Intelligence Conference 2017;11.-12 September 2017 Duisburg, Germany(4):1817–39.

Hecking T, Steinert L, Masías VH, Hoppe HU. Relational patterns in cross-media information diffusion networks. In: International Workshop on Complex Networks and their Applications VI. Springer; 2018. p. 1002–14. doi:10.1007/978-3-319-72150-7.

Henderson M, Johnson NF, Auld G. Silences of ethical practice: Dilemmas for researchers using social media. Educational Research and Evaluation 2013;19(6):546–60. doi:10.1080/13803611.2013.805656.

House F. Freedom of the Press 2015: Harsh laws and violence drive global decline. Freedom House, 2015.

Hu T, Zhang P. Social media usage as a formative construct: Conceptualization, validation, and implication. Journal of Information Technology Management 2016;27(4):151–68.

Husson F, Josse J. Visualization and Verbalization of Data; CRC Press Boca Raton, FL. p. 165–84.

Husson F, Lê S, Pagès J. Exploratory multivariate analysis by example using R. CRC press, 2017.

Johnsson M, editor. Applications of Self-Organizing Maps. InTech, 2012. doi:10.5772/3464.

Josse J, Chavent M, Liquet B, Husson F. Handling missing values with regularized iterative multiple correspondence analysis. Journal of Classification 2012;29(1):91–116. doi:10.1007/s00357-012-9097-0.

Josse J, Husson F, et al. missmda: a package for handling missing values in multivariate data analysis. Journal of Statistical Software 2016;70(1):1–31.

Kaslow FW, Patterson T, Gottlieb M. Ethical dilemmas in psychologists accessing internet data: Is it justified? Professional Psychology: Research and Practice 2011;42(2):105–12. doi:10.1037/a0022002.

Kim Y, Chen HT. Social media and online political participation: The mediating role of exposure to cross-cutting and like-minded perspectives. Telematics and Informatics 2016;33(2):320–30. doi:10.1016/j.tele.2015.08.008.

Kruikemeier S, Shehata A. News media use and political engagement among adolescents: An analysis of virtuous circles using panel data. Political Communication 2016;34(2):221–42. doi:10.1080/10584609.2016.1174760.

Kuhn M, Weston S, Coulter N, Quinlan R. C50: C5. 0 decision trees and rule-based models. R package version 01 0-21, URL http://CRAN R-project org/package C 2014;50.

Lebart L, Salem A, Berry L. Visualization of Textual Data. Springer Netherlands, 1998. doi:10.1007/978-94-017-1525-6_6.

Miyamoto S, Ichihashi H, Honda K. Algorithms for Fuzzy Clustering. Springer Berlin Heidelberg, 2008. doi:10.1007/978-3-540-78737-2.

Mourão R, Saldaña M, McGregor S, Zeh A. Support for protests in Latin America: Classifications and the role of online networking. Social Sciences 2016;5(4):58. doi:10.3390/socsci5040058.

Opp KD. Theories of political protest and social movements: A multidisciplinary introduction, critique, and synthesis. Routledge, 2009. URL: https://doi.org/10.4324%2F9780203883846. doi:10.4324/9780203883846.

Pérez-Liñán A, Polga-Hecimovich J. Explaining military coups and impeachments in Latin America. Democratization 2016;24(5):839–58. doi:10.1080/13510347.2016.1251905.

Qiu W, Joe H. clustergeneration: Random cluster generation. R package version 2009;1(7).

Quinlan J. Data mining tools see5 and c5. 0. st. ives nsw, australia: Rulequest research. 2003.

Quinlan J. C4.5: Programs for Machine Learning. Ebrary online. Elsevier Science, 2014. URL: https://books.google.de/books?id=b3ujBQAAQBAJ.

Quinlan JR. Induction of decision trees. Machine Learning 1986;1(1):81–106. doi:10.1007/bf00116251.

Renaud-Gentié C, Burgos S, Benoît M. Choosing the most representative technical management routes within diverse management practices: Application to vineyards in the loire valley for environmental and quality assessment. European Journal of Agronomy 2014;56:19–36. doi:10.1016/j.eja.2014.03.002.

Ruijgrok K. From the web to the streets: Internet and protests under authoritarian regimes. Democratization 2016;:1–23doi:10.1080/13510347.2016.1223630.

Saldana M, Mcgregor SC, Gil De Zuniga H. Social media as a public space for politics: Cross-national comparison of news consumption and participatory behaviors in the United States and the United Kingdom. International Journal of Communication 2015;9(1):3304–26. URL: http://ijoc.org/index.php/ijoc/article/view/3238/1485.

Skoric MM, Poor ND, Liao Y, Tang SWH. Online organization of an offline protest: From social to traditional media and back. In: 2011 44th Hawaii International Conference on System Sciences. IEEE; 2011. doi:10.1109/hicss.2011.330.

Skoric MM, Zhu Q, Goh D, Pang N. Social media and citizen engagement: A meta-analytic review. New Media & Society 2016a;18(9):1817–39. doi:10.1177/1461444815616221.

Skoric MM, Zhu Q, Pang N. Social media, political expression, and participation in Confucian Asia. Chinese Journal of Communication 2016b;9(4):331–47. doi:10.1080/17544750.2016.1143378.

Steinert-Threlkeld ZC, Mocanu D, Vespignani A, Fowler J. Online social networks and offline protest. EPJ Data Science 2015;4(1). doi:10.1140/epjds/s13688-015-0056-y.

Towner TL. All political participation is socially networked? Social Science Computer Review 2013;31(5):527–41. doi:10.1177/0894439313489656.

Tukey JW. Exploratory data analysis. volume 2. Reading, Mass., 1977.

Valenzuela S. Unpacking the use of social media for protest behavior: The roles of information, opinion expression, and activism. American Behavioral Scientist 2013;57(7):920–42. doi:https://doi.org/10.1177/0002764213479375.

Valenzuela S, Somma NM, Scherman A, Arriagada A. Social media in Latin America: Deepening or bridging gaps in protest participation? Online Information Review 2016;40(5):695–711. doi:10.1108/oir-11-2015-0347.

Waheed H, Anjum M, Rehman M, Khawaja A. Investigation of user behavior on social networking sites. PLOS ONE 2017;12(2):1–19. doi:10.1371/journal.pone.0169693.

Ward JH. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 1963;58(301):236–44. doi:10.2307/2282967.

Wolfsfeld G, Yarchi M, Samuel-Azran T. Political information repertoires and political participation. New Media & Society 2016;18(9):2096–115. doi:10.1177/1461444815580413.