ASSOCIATION FOR COMMUNICATION EXCELLENCE

Journal of Applied Communications

Volume 106 | Issue 3 Article 3

A Statistical Approach to Classification: A guide to hierarchical cluster analysis in agricultural communications research

Ch'Ree Essary
Texas Tech University

Laura Morgan Fischer
Texas Tech University

Erica Irlbeck
Texas Tech University

Recommended Citation

Essary, Ch'Ree; Fischer, Laura Morgan; and Irlbeck, Erica (2022) "A Statistical Approach to Classification: A guide to hierarchical cluster analysis in agricultural communications research," *Journal of Applied Communications*: Vol. 106: Iss. 3. https://doi.org/10.4148/1051-0834.2431

A Statistical Approach to Classification: A guide to hierarchical cluster analysis in agricultural communications research

Abstract

Classification, the sorting of similar objects or organisms into groups based on shared qualities and characteristics, is how we make sense of the world. As the field of agricultural communication and our understanding of media effects becomes more complex, it is important to have approaches that allow for a valid and reliable method of classifying units of analysis — whether they are texts, people, or other artifacts — into groups based on theoretically sound variables. This paper discusses one method of classification, the hierarchical cluster analysis, and how this method may be applied by 1) Developing Variables for Study, 2) Choosing a Sample, 3) Removing Unnecessary Variables, 4) Running the analysis, and 5) Interpreting Clusters. This professional development paper suggests this method could have positive implications for agricultural and science communication research including increased validity and reliability, rigorous development, and deeper understanding of mass communication theory. In addition, we provide recommendations for future research such as audience segmentation in agricultural and science communication research.

Keywords

Hierarchical cluster analysis, audience segmentation, framing analysis, data analysis

Cover Page Footnote/Acknowledgements

This paper was presented as an innovative idea poster and as a professional development session at the 2022 National Agricultural Communications Symposium.

Humans have the natural, visceral instinct to classify. Classification, defined as the sorting of similar objects or organisms into groups based on shared qualities and characteristics, is how we make sense of the world (Batley, 2014). Throughout our early education, we were taught to classify in order to group similar objects together and expand our understanding of the world. For example, in science class, we were taught to classify living things into Domain, Class, Genus, Family, and when we went to the library, we were taught to classify books using the Dewey Decimal System. Building on the work of early philosophers Aristotle and Theophrastos, Carl Linnaeus created the foundation for our taxonomy of living things, putting plants in groups based on their characteristics. This early scientific researcher, (Linnaeus, 1737, as cited in Everitt et al., 2011) explains the use of classification:

> All the real knowledge which we possess, depends on methods by which we distinguish the similar from the dissimilar. The greater the number of natural distinctions this method comprehends the clearer becomes our idea of things. The more numerous the objects which employ our attention the more difficult it becomes to form such a method and the more necessary.

Hierarchical cluster analysis (HCA) is a method of statistical analysis that is used to develop a set of nested categories or clusters, which are created by sequentially pairing variables (Bridges, 1966). This method of analysis identifies homogeneous groups of cases based on pre-determined characteristics (Everitt et al., 2011). HCA is no new concept; cluster analysis is a common tool in the field of marketing, specifically for creating consumer segmentation (Cibulková & Sulc, 2018). However, HCA has become a useful method of deriving frames of communication in science communication (i.e. Donk et al., 2011; Matthes & Kohring, 2008) and the larger field of mass communication (i.e. David et al., 2011; Guenther et al., 2020). More recently, this method has been used to identify clusters of audiences for science communication (i.e. Runge et al., 2018), which has been identified as a growing need in the field (Füchslin, 2019; Hine et al., 2014; Maibach et al., 2011; Moser, 2014). Though it's use has increased in mass communication, it has only been used once in the more applied field of agricultural communications (i.e. Steede et al., 2018).

Though HCA has been growing in popularity in the larger field of mass communication and science communication, there is quite a learning curve to conducting this analysis as it is not a method commonly taught in data analysis courses. This professional development paper seeks to address this need by providing a discussion of how to use HCA. In the methods and procedures section, we will outline the steps to complete an HCA using an agricultural communication example from study design and data collection to analysis and interpretation. This paper will then discuss theoretical implications, potential future applications, and advantages of HCA in agricultural communication research.

Methods and Procedures

For the purposes of this paper, an empirical study with the purpose of identifying frames in news articles regarding genetic modification in agriculture through a quantitative content analysis will be used as a demonstration of HCA in practice. Reliability was established in the codebook utilized in this content analysis using three coders; intercoder reliability was calculated using Krippendorf's alpha reliability coefficient for each variable ($\alpha > 0.85$). Though HCA was used as a way of conducting framing analysis in this example, it should be noted that this analysis can also be used to analyze data collected using other quantitative methods, such as survey data. Dichotomous measures were used in this analysis. However, continuous measures can be used, as well (see Runge et al., 2018 for an example).

The example shown analyzes data using SPSS Version 27 (2021). To provide the exact analyses ran and the more clarification on the steps of the analysis, SPSS syntax with a commentary of each decision can be found at this Open Science Framework page: https://osf.io/anvg4/?view_only=dbf8be3a0e944e2d823734a6347cf595. The data used in this example can also be found on this OSF page for those readers who would like to practice conducting the analysis themselves.

1. *Developing Variables for Study*

As HCA clusters cases by chosen variables, characteristics of the item of analysis need to be chosen. The study at hand sought to identify frames in news coverage, using framing theory was used as a framework for this study. This study conceptualized frames as themes, actors, perceived risks and benefits, and tone of the article chosen by the journalist and editor; these elements cluster together in systematic ways that create the frame. Thus, the central variables, or the frame elements, chosen for analysis in this study comprised of *Theme*, *Actor*, *Risks/Costs Mentioned*, *Benefits Mentioned*, and *Overall Tone.* Within each of these variables, sub-variables were developed to account for the various themes, actors, and the risks and benefits mentioned in the article as well as the overall sentiment. In total, there were 25 sub-variables coded in this study. Data were coded dichotomously (1 = *present*, 0 = *not present*). Selected variables and their descriptions are displayed in Table 1.

When identifying clusters of survey participants, in audience segmentation, say, the variables are often based on geographic location, personality traits, demographics, differences in the use of products, and psychographics (Goyat, 2011). An audience segmentation analysis may use variables such as attitudes, behaviors, beliefs, values, and knowledge of the subject at hand among others (Hine et al., 2014; National Academies of Sciences, Engineering, and Medicine [NASEM], 2017; Schäfer et al., 2018). Runge and colleagues (2018), for example, used participants' deference to science authority, religiosity, and political ideology as segmentation variables, for example. Variables should be chosen based on findings of previous literature and theoretical paradigms and models.

DOI: 10.4148/1051-0834.2431

Table 1

Selected variables for analysis

Frame Element	Variables	Description
Theme	Economic	Article deals with the economic opportunities, risks and costs of GM.
	Research	Article deals with scientific research conducted concerning GM.
	Public Opinion	Article mentions public opinion surveys or public protests or demonstrations.
	Policy Change	Article mentions policy change. This could be state or federal, policy changes in a business, or mandatory or voluntary labeling requirements.
	Health	Article deals with health concerns or benefits of GM.
	Environment	Article deals with environmental concerns or benefits of GM.
	International	Article deals with trade and food production in other countries.
	Judicial	Article deals with a judicial case regarding GM.
Actor	Politicians	Members of Congress, representatives, senators, the president, etc.
	Interest groups	Advocacy groups such as the Grocery Manufacturers Association, Farm Bureau, Just Label It, etc.
	Restaurants/Grocery Stores	Dine-in and fast-food restaurants and grocery stores such as Chipotle, Whole Foods, etc.
	Chemical and Ag-Science Companies	Chemical and agriculture companies such as Monsanto, Bayer, DuPont, etc.
	Food Manufacturers	Companies that make food products such as Nestle, General Mills, Hershey, CHS, etc.
	Producers	Agriculture producers such as farmers and ranchers — organic or non-organic.
	Consumers	The general public, non-producers, who buy food products.
	Federal or State Department or Agency	Departments and agencies such as USDA, FDA, EPA, including the Secretary of Agriculture.
	Scientists	Scientists at research institutions, universities, private corporations, government agencies.
Risks/Cost	Health	Identifies GM as a risk for human or animal health.
	Producers	Identifies GM as a risk/cost for producers.
	Consumer	Identifies GM as a risk/cost for consumers.
	Environmental	Identifies GM as a risk/cost for the environment.
	Pesticide Use	Identifies the usage of pesticides as a negative cost of GM.

(Continued)

Frame Element	Variables	Description
Benefits	Health	Identifies GM as a benefit for health.
	Producers	Identifies GM as a benefit cost for producers.
	Consumer	Identifies GM as a benefit for consumers.
	Environmental	Identifies GM as a benefit for the environment.
Tone	Positive	The article is pro-genetic modification.
	Negative	The article is anti-genetic modification.
	Neutral	The article is balanced in reporting.

2. Choosing your sample

The data used for this example was originally collected through a content analysis of news frames regarding genetic modification in agriculture during a time in which GMOs were a highly politicized science issue. This study employed a census, using the Nexis Uni Database. One hundred articles were selected using search terms, "Genetically Modified Organisms," "G.M.O.," "GMO," "Genetic modification," "genetically modified food," "GM food," or "National Bioengineered Food Disclosure" in the headline and lead sections in *The New York Times*, *USA Today*, and *The Associated Press* between January 1, 2015, and December 31, 2020.

Though HCA is commonly used in other fields (Cibulková & Sulc, 2018), there is no standard rule-of-thumb for sample size in hierarchical cluster analysis (Sarstedt & Mooi, 2011). Dolnicar (2002) recommended a minimum sample size of 2^k (k = number of variables for analysis), but also a preferred sample size of $5*2^k$. In contrast, Breckenridge (2000) proposed a sample of at least 120 cases. Dolnicar (2002) found that analyses with sample sizes as small as 10 cases and as large as 20,000 cases have been used in previous literature. More recent cluster analyses have used 270 (see Steede et al. 2019), 709 (see Di Vita et al. 2021), and 179 (see Bejaei, Cliff, & Singh, 2020), just to name a few. In the example data, a sample of N = 100, which was a census of the population, was used.

3. Removing Unnecessary Variables

After collecting data from all cases, whether they be survey respondents or news articles, it is possible that variables chosen for analysis were only present in a few of the articles. These variables, therefore, do not contribute to the cluster. As hierarchical cluster analysis is sensitive to outliers (Everitt et al., 2011), it is necessary to remove variables that do not contribute to the cluster or grouping. Matthes and Kohring (2008) recommend removing those variables which were present in less than 5% of the cases. In the example shown, sub-variables *Theme, Economics*; *Theme, Environment*; and *Risk*, *Producers* were removed as they were only present in n = 3, n = 3, and n = 2 of the cases, respectively. This step is not necessary if continuous data is used, if one were using HCA to identify audience segments (see Runge et al., 2018 for an example).

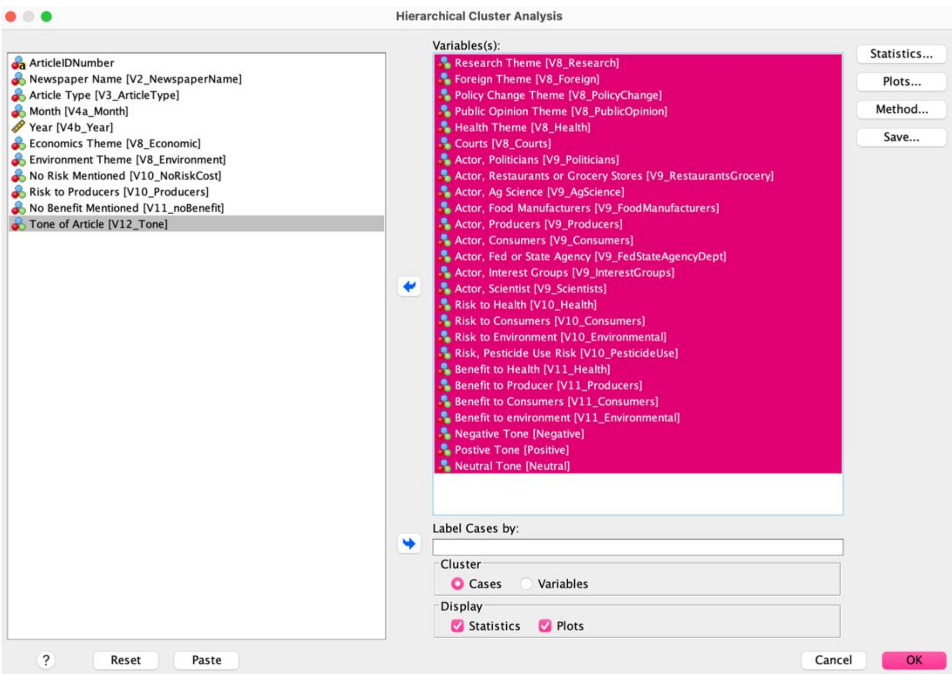
4. Running the Analysis

In this example, researchers chose to run the HCA using Ward's Method. Ward's Method, also called minimum sum of squares clustering, is a method of hierarchical clustering in which the agglomeration of two cases into a cluster is based on the size of an error sum-of-squares criterion. At each stage of clustering the goal of Ward's Method is to minimize the increase in the total within-cluster error sum of squares. In other words, the distance between clusters is defined as an increase in sums of squares within the clusters and then, after fusion of the cases, summed over all variables (Everitt et al., 2011; Ward, 1963). Ward's Method tends to create clusters that are around the same size, but as mentioned earlier, is very susceptible to outliers (Everitt et al., 2011). Thus, unnecessary variables need to be omitted from the inclusion in the HCA, as discussed in step 3.

Identification of the Appropriate Number of Clusters. There are multiple steps to running the HCA: first, an initial hierarchical cluster analysis using Ward's method, squared Euclidean distance is run to identify the appropriate number of clusters. The HCA command can be found in **Analyze>Classify>Hierarchical Cluster**. Select the variables which will be used for analysis, remembering to not include those variables previously determined unnecessary. Select "Cluster by Cases" and ask for statistics and plots as shown in Figure 1.

Figure 1

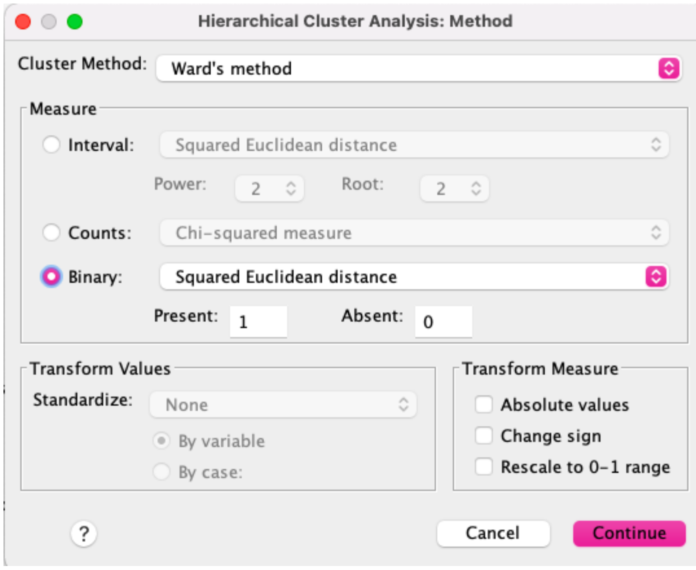
Hierarchical cluster analysis main screen SPSS Version 27 (2021)



Next, in the Statistics menu, ask for an agglomeration schedule. In the Plots menu, ask for a dendrogram. And finally, in the Method menu, select Ward's method, binary measure, and Squared Euclidean Distance. This is displayed in Figure 2. The analysis is now ready to be run. Select "OK."

Figure 2

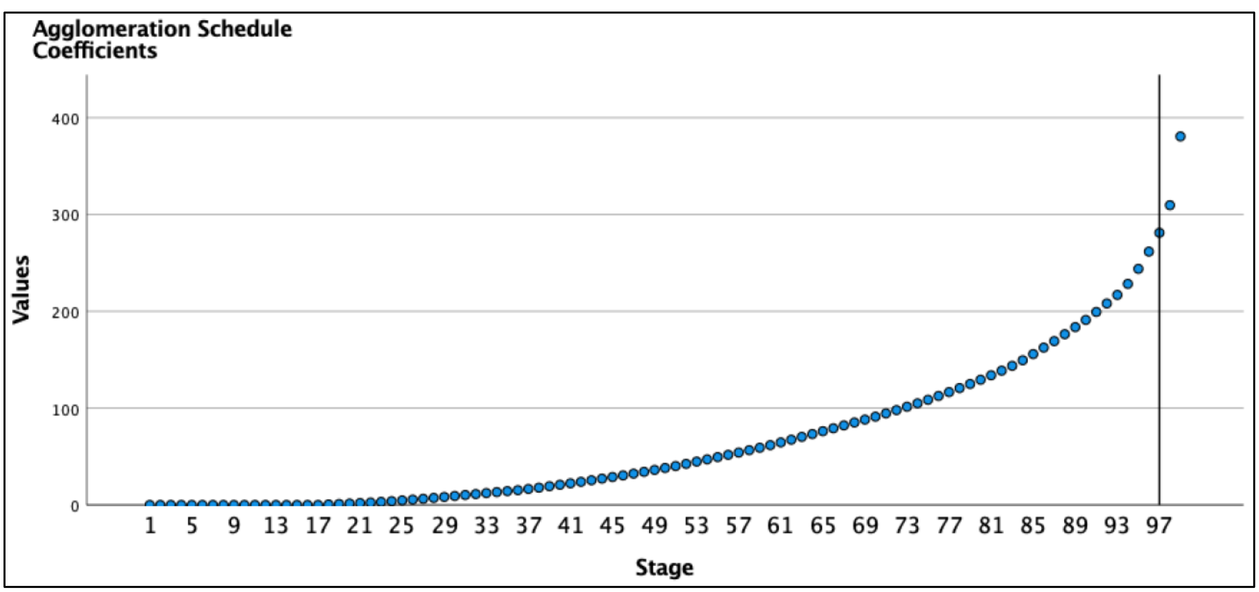
Hierarchical cluster analysis method screen



The number of appropriate clusters can be found by analyzing large breaks in the coefficients of the agglomeration schedule. Agglomeration refers to how the sample is massed and at which stage in clustering the cases' differences require them to be assigned to different clusters. The simplest way to assess where this gap is by viewing the coefficients in a line graph, similar to a scree chart. In the example output shown in Figure 3, we can see a break at the 97th stage of clustering.

Figure 3

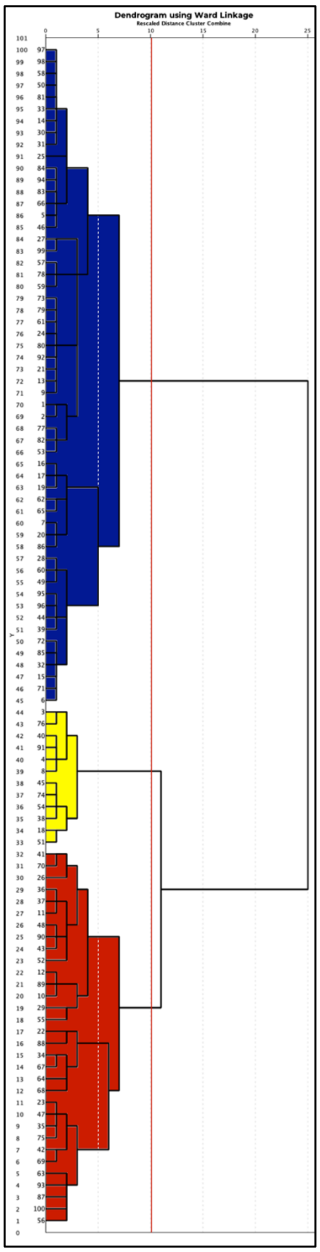
Agglomeration Schedule Coefficient Line Graph



To determine the appropriate number of clusters, you will take the sample size number and then subtract the number at the stage of clustering break. In our case, we used our sample size of, 100, and subtracted the stage of clustering break of 97. Thus, our sample can be broken into three clear clusters. This can be verified through a visual analysis of a dendrogram of the clusters. When using continuous measures, a Levene's test (homogeneity of variance) can be used to verify the clusters as seen in Runge et al. (2018). The dendrogram resulting from the example data is displayed in Figure 4.

Figure 4

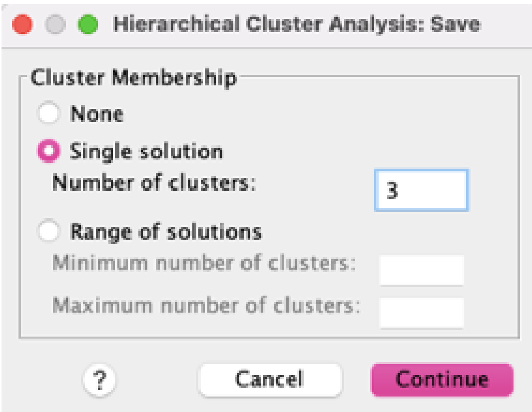
Dendrogram of clusters



Assigning Member Cases to Each Cluster. After the identification of clusters and this confirmation, a second HCA is conducted, this time indicating there will be three clusters so that SPSS may assign member cases to each cluster. To indicate the number of clusters, select "Single solution" in the Save menu and input the number of clusters (shown in Figure 5). SPSS automatically creates a new column for cluster membership, assigning a number for each cluster.

Figure 5

Saving cluster membership SPSS Version 27 (2021)



5. Interpreting Clusters

Classifying a sample into groups without knowing what characteristics define each group is inefficacious, to say the least. Would having classes of mammals and reptiles be useful if we didn't know the characteristics which set us apart? Warm-blooded versus cold-blooded, fur versus scales. Not at all. Thus, the next step in hierarchical clustering is determining which of our selected variables are the most present in each cluster. To do so, we will compare the means of the selected variables and clusters. As displayed in Table 2, those variables with the highest means in each cluster are those that define the cluster. For example, the theme of research had a mean of .00 for Cluster 1, .08 for Cluster 2, and .24 for Cluster 3. Because .24 was the highest mean, we know the theme of research belongs to Cluster 3. Previous studies that have used this method (i.e. Matthes & Kohring, 2008) did not note standard deviation has having an impact on how clusters are interpreted. However, they should be reported as well per APA 7 statistical reporting requirements (American Psychological Association, 2019).

https://newprairiepress.org/jac/vol106/iss3/3
DOI: 10.4148/1051-0834.2431

Table 2

Mean Values and Standard Deviations of Variables on Clusters

Frame Elements	Cluster 1 (*n* = 56) *M (SD)*		Cluster 2 (*n* = 12) *M (SD)*		Cluster 3 (*n* = 32) *M (SD)*	
Theme, Research	.00	(.00)	.08	(.289)	**.24**	**(.429)**
Theme, Public Opinion	.14	(.353)	**1.00**	**(.00)**	.41	(.499)
Theme, Policy Change	**.96**	**(.187)**	.58	(.515)	.28	(.457)
Theme, Health	.00	(.000)	.08	(.289)	**.41**	**(.499)**
Theme, International	**.11**	**(.312)**	.00	(.00)	.06	(.246)
Theme, Judicial	**.13**	**(.334)**	.00	(.00)	.00	(.00)
Actor, Politicians	**.64**	**(.483)**	.25	(.452)	.09	(.296)
Actor, Interest Groups	**.45**	**(.502)**	.00	(.000)	.13	(.336)
Actor, Restaurants & Grocery Stores	**.16**	**(.371)**	.08	(.289)	.00	(.00)
Actor, Ag Science Companies	.16	(.371)	.08	(.289)	**.28**	**(.457)**
Actor, Food Manufacturer	**.41**	**(.496)**	.17	(.389)	.06	(.246)
Actor, Producers	.07	(.260)	.08	(.289)	**.25**	**(.440)**
Actor, Consumers	.25	(.437)	**1.00**	**(.000)**	.72	(.457)
Actor, Government Agency or Department	.21	(.414)	.17	(.389)	**.28**	**(.457)**
Actor, Scientist	.04	(.187)	.08	(.289)	**.53**	**(.507)**
Risk, Health	.11	(.312)	**.42**	**(.515)**	.19	(.397)
Risk, to Consumers	.02	(.134)	.08	(.12)	**.09**	**(.296)**
Risk, to Environment	.05	(.227)	**.25**	**(.452)**	.13	(.336)
Risk, Pesticide Use	.04	(.187)	**.42**	**(.515)**	.13	(.336)
Benefit, to Health	.05	(.227)	.00	(.000)	**.34**	**(.483)**
Benefit, to Producer	.02	(.134)	.00	(.000)	**.53**	**(.507)**
Benefit, to Consumers	.04	(.187)	.00	(.000)	**.47**	**(.507)**
Benefit, to Environment	**.09**	**(.288)**	.00	(.000)	.00	(.000)
Tone, Positive	.00	(.000)	.00	(.00)	**.53**	**(.507)**
Tone, Negative	.02	(.134)	**.92**	**(.289)**	.06	(.246)
Tone, Neutral	**.98**	**(.134)**	.08	(.289)	.41	(.499)

The examination of the means of each frame element on the three clusters as displayed in Table 2 tells us that *Themes Policy Change*, *International*, and *Judicial*, *Actors*, *Politicians*, *Interest Groups*, *Restaurants & Grocery Stores*, and *Food Manufacturers*, *Benefit to Environment*, and *Tone*, *Neutral* are most prominent in Cluster 1. In Cluster 2, *Theme*, *Public Opinion*, *Actor*, *Consumers*, *Risks*, *to Health*, *to Environment*, *of Pesticide Use*, and *Tone*, *Negative* are most prominent. Finally, *Themes, Research* and *Health*, *Actors. Ag Science Companies*, *Producers*, *Government Agencies or Departments*, *Risk*, *to Consumers*, Benefits, *to Health*, *to Producers*, and *to Consumers*, and *Tone, Positive* were most prominent. By knowing which elements were most prominent in each frame, we can now name our clusters. For example, Cluster 1 could be named Policy Change Conflicts as stories within this cluster highlight politicians and other political actors debating legislation regarding genetically modified organisms.

Discussion and Conclusion

HCA is an underused method of study design and statistical analysis in agricultural communication research that could have an abundantly positive impact on the field. As it relates to content analysis, shown in the example provided, this computerized statistical method of inferring frames provides a robust understanding of journalistic practice and the formation of frames by coding for individual indicators that characterize a frame. This method can be used to explore untapped areas in agricultural communication research which can improve the way scholars in this field tailor messages to audiences.

These conceptualizations of framing theory support a statistical analysis such as HCA to determine how these different frame elements which make up a frame cluster together as Entman (1993) suggested. Operationalizing this conceptual definition, as first suggested by Matthes and Kohring (2008), can prove useful in future framing analyses. The current example used in this paper identified three frames used by national news publications when covering genetic modification in agriculture providing a nuanced way of how certain elements of a news story cluster together to frame an issue in a certain light. This method of framing analysis using HCA has been used once in agricultural communications research by Steede et al. (2018) to identify news frames used by mainstream news publications when covering antibiotic use in livestock. However, it should be used more in order to provide a more nuanced understanding of news frames. This paper, hopefully, contributes to the agricultural communications field by providing a step-by-step tutorial of how to run this analysis with example data.

Although this paper provided an example of HCA in the context of a framing analysis, the methods in this paper can also be used to tap into an area of research currently unexplored in the current agricultural communications literature: audience segmentation. Audience segmentation is a growing area in science communication research (Füchslin, 2019; Hine et al., 2014; Maibach et al., 2011; Moser, 2014) in which "a large potential audience into subgroups and tailoring messages differently for each subgroup" (NASEM, 2017, p. 56). By segmenting audiences, agricultural science communication scholars can better tailor messages for different audiences, avoiding directly challenging pre-existing, strongly held beliefs while still providing accurate information (NASEM, 2017). HCA could be used to measure those individual schemata that guide message processing to create audience segments for science and agricultural communication potentially allowing academics and practitioners alike to effectively craft effective messaged to lay audiences. An example of this, mentioned throughout this paper is Runge et al. (2018). Runge and colleagues (2018) used HCA to identify audiences for science communication by segmenting a general population into segments based on their deference to science authority, political ideology, and religiosity. Similar studies have been conducted in recent years in the science communication literature (see Füchslin, 2019). However, this area has yet to be tapped into in the applied field of agricultural communications.

This paper sought to provide a step-by-step tutorial of how to conduct HCA in agricultural communications research, outlining theoretical relevancy of the analysis method and proper procedures that should be followed. It also outlined recommendations for future research using this method. Additional resources for HCA can be found below:

- Textbook on Cluster Analysis
    - Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Wiley.
- Original Example of HCA used for Content Analysis
    - Matthes, J., & Kohring, M.. (2008). The content analysis of media frames: Toward improving reliability and validity. *Journal of Communication, 58*(2), 258–279. https://doi.org/10.1111/j.1460-2466.2008.00384.x
- Example of HCA used for Audience Segmentation (continuous variables)
    - Runge, K.K., Brossard, D., & Xenos, M.A. (2018). Protective progressives to distrustful traditionalists: A post hoc segmentation method for science communication. *Environmental Communication*, 12, 1023–1045. https://doi.org/10.1080/17524032.2018.1513854

References

American Psychological Association (2019). *Publication manual of the American Psychological Association*. American Psychological Association.

Batley, S. (2014). *Classification in theory and practice*. Chandos Publishing. https://doi.org/10.1080/14649055.2005.10766077

Bejaei, M., Cliff, M. A., & Singh, A. (2020). Multiple correspondence and hierarchical cluster analyses for the profiling of Fresh Apple customers using data from two marketplaces. *Foods, 9*(7), 873. https://doi.org/10.3390/foods9070873

Breckenridge, J. N. (2000). Validating cluster analysis: Consistent replication and symmetry. *Multivariate Behavioral Research, 35*(2), 261–285. https://doi.org/10.1207/S15327906MBR3502_5

Bridges, C. C. (1966). Hierarchical cluster analysis. *Psychological Reports, 18*(3), 851–854. https://doi.org/10.2466/pr0.1966.18.3.851

Cibulková, J., & Sulc, Z. (2018). *A case study of customer segmentation with the use of hierarchical cluster analysis of categorical data.* Proceedings of the Applications of Mathematics and Statistics in Economics, AMSE, Kutná Hora, Czech Republic. http://www.amse-conference.eu/wp-content/uploads/2018/09/AMSE_2018_Cibulkova_Sulc.pdf

David, C. C., Atun, J. M., Fille, E., & Monterola, C.. (2011). Finding frames: Comparing two methods of frame analysis. *Communication Methods and Measures*, *5*(4), 329–351. https://doi.org/10.1080/19312458.2011.624873

Di Vita, G., Zanchini, R., Falcone, G., D’Amico, M., Brun, F., & Gulisano, G. (2021). Local, Organic or Protected? Detecting the role of different quality signals among Italian olive oil consumers through a hierarchical cluster analysis. *Journal of Cleaner Production, 290*(125795). https://doi.org/10.1016/j.jclepro.2021.125795

Dolnicar, S. (2002). *A review of unquestioned standards in using cluster analysis for data driven market segmentation*. Paper presented at the Australian and New Zealand Marketing Academy Conference 2002, Melbourne, Australia. https://core.ac.uk/download/pdf/36990441.pdf

Donk, A., Metag, J., Kohring, M., & Marcinkowski, F. (2012). Framing emerging technologies: Risk perceptions of nanotechnology in the German press. *Science Communication, 34*(1), 5–29. https://doi.org/10.1177/1075547011417892

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication, 43*(4), 51-58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Wiley.

Füchslin, T. (2019). Science communication scholars use more and more segmentation analyses: can we take them to the next level? *Public Understanding of Science, 28*(7), 854-864. *doi/10.1177/0963662519850086*

Goyat, S. (2011). The basis of market segmentation: a critical review of literature. *European Journal of Business and Management*, *3*(9), 45-54. https://core.ac.uk/download/pdf/234624114.pdf

Guenther, L., Ruhrmann, G., Bischoff, J., Penzel, T., & Weber, A.. (2020). Strategic framing and social media engagement: Analyzing memes posted by the German identarian movement on Facebook. *Social Media + Society*, *6*(1), 205630511989877. https://doi.org/10.1177/2056305119898777

Hine, D. W., Reser, J. P., Morrison, M., Phillips, W. J., Nunn, P., & Cooksey, R. (2014). Audience segmentation and climate change communication: Conceptual and methodological considerations. *Wires Climate Change*, *5*(4), 441–459. https://doi.org/10.1002/wcc.279

Matthes, J., & Kohring, M.. (2008). The content analysis of media frames: Toward improving reliability and validity. *Journal of Communication*, *58*(2), 258–279. https://doi.org/10.1111/j.1460-2466.2008.00384.x

Maibach, E. W., Leiserowitz, A., Roser-Renouf, C., & Mertz, C. K. (2011). Identifying like-minded audiences for global warming public engagement campaigns: An audience segmentation analysis and tool development. *PLoS ONE*, *6*(3), e17571. https://doi.org/10.1371/journal.pone.0017571

Moser, S. C. (2014). Communicating adaptation to climate change: The art and science of public engagement when climate change comes home. *Wires Climate Change, 5*(3), 337–358. https://doi.org/10.1002/wcc.276

National Academies of Sciences, Engineering, & Medicine. (2017). *Communicating science effectively: A research agenda.* National Academies Press. https://doi.org/10.17226/23674

Runge, K.K., Brossard, D., & Xenos, M.A. (2018). Protective progressives to distrustful traditionalists: A post hoc segmentation method for science communication.

Environmental Communication, 12, 1023–1045. https://doi.org/10.1080/17524032.2018.1513854

Sarstedt, M., & Mooi, E. (2011). *A concise guide to market research: The process, data, and methods using IBM SPSS statistics* (2nd ed.). Springer. https://doi.org/10.1007/978-3-642-12541-6

Schäfer, M. S., Füchslin, T., Metag, J., Kristiansen, S., & Rauchfleisch, A. (2018). The different audiences of science communication: A segmentation analysis of the Swiss population's perceptions of science and their information and media use patterns. *Public Understanding of Science*, *27*(7), 836–856. https://doi.org/10.1177/0963662517752886

Silk, K. J., Weiner, J., & Parrott, R. L.. (2005). Gene cuisine or Frankenfood? The theory of reasoned action as an audience segmentation strategy for messages about genetically modified foods. *Journal of Health Communication, 10*(8), 751–767. https://doi.org/10.1080/10810730500326740

Steede, G. M., Meyers, C., Li, N., Irlbeck, E., & Gearhart, S. (2019). A content analysis of antibiotic use in livestock in national U.S. newspapers. *Journal of Applied Communications, 103*(1). https://doi.org/10.4148/1051-0834.2237

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*(301), 236-244. https://doi.org/10.1080/01621459.1963.10500845