

Attribute agenda setting and information overload

Computer-assisted analysis for understanding compelling arguments

Magdalena Saldaña
Texas Tech University

The compelling argument concept, one of the least studied components of attribute agenda setting, suggests that some attributes increase the salience of an object on the public agenda of issues. By conducting two studies, this article examines the compelling argument concept applying both manual content analysis (Study 1) and computerized-analysis tools (Study 2), considering frequency and degree centrality as measures of attribute salience. Results show that the application of computer-aided methods and mathematical techniques can efficiently identify attributes and estimate degree centrality, which are the core elements of the second and third level of the agenda-setting theory, respectively. Also, findings indicate that absolute frequency, rather than the presence or absence of an attribute in a news story, is a more parsimonious measure of redundancy to identify compelling arguments in news stories. This study proposes methodological innovations that further expand the scope of attribute agenda setting in the big data landscape.

Theoretically, the core assertion in traditional agenda-setting research is that news media attention to specific “objects” (e.g., issues, candidates, companies, activist groups) leads to increased public concern toward those same objects. Further studies have shown that not only the objects covered by the media, but how such objects are portrayed, affect the public salience of both the object and its attributes. This is the second level of the agenda-setting theory, also known as *attribute agenda setting*: “the attributes emphasized in news coverage become more salient in the minds of media consumers and more influential in terms of actual effects on opinions and attitudes” (Bowe, Fahmy, & Wanta, 2013, p. 637).

Methodologically, researchers have examined news media attention by content-analyzing news stories from different outlets and about numerous issues. Berelson (1952) defined content analysis as “a research technique for the objective,

systematic, and quantitative description of the manifest content of communication” (p.18). Once the sample of text corpuses is selected and the categories are defined for coding, the following step is training coders and assessing reliability (McMillan 2000; Riffe, Lacy & Fico, 2005). With more than 400 studies and counting, many of the agenda-setting studies have followed this method to analyze news media content. What happens when the volume of information exceeds the capability of human coders? In an era of “big data,” how realistic are the expectations of analyzing news content manually? The available evidence indicates these expectations are hardly realistic.

By conducting two different studies, this article poses two innovations to further expand the scope of attribute agenda setting in a context of information overload. First, I propose a new way to identify and measure attributes in the news media agenda, emphasizing the role of redundancy. Second, I incorporate computerized content analysis and mathematical techniques to assess attributes’ effects on public opinion. This article adds to and improves upon the available methods for studying agenda setting.

Attribute agenda setting and compelling arguments

The second level of the agenda setting poses that the salience of the objects’ attributes emphasized by the media is also transferred to the public (McCombs, Llamas, Lopez-Escobar & Rey, 1997). Ghanem (1996, 1997) noticed that some attributes were more likely than others to be included in media messages, and more likely to be remembered by the public. By correlating the salience of individual attributes on the media agenda with the salience of the object on the public agenda, Ghanem found that some attributes had a diagonal effect that, in some cases, was even stronger than the direct effect at the first level. In her study about crime coverage in Texas, Ghanem found that stories describing situations where people would feel personally threatened, or stories describing a crime that occurred in Texas, had a stronger effect than the actual amount of crime coverage. These attributes are examples of *compelling arguments*, and they provide another path for the transfer of salience between the media agenda and the public agenda (Kioussis, 2005; McCombs, 2004, 2014; McCombs & Ghanem, 2001; Yioutas & Segvic, 2003).

The extensive research conducted by agenda-setting scholars has contributed substantial empirical evidence of agenda-setting effects. However, many of the studies in agenda-setting research focus on first-level agenda-setting effects (Kim & Kim, 2014). Relatively few empirical studies have looked into the compelling argument concept, considered by Sheafer (2007) as “a rather neglected part of second-level agenda setting” (p. 22).

On the media agenda, object or attribute salience usually refers to the coverage of certain elements over time (Chyi & McCombs, 2004). The concept is usually measured by the frequency of an object or attribute appearing in the news coverage. Most studies consider frequency as presence/absence of the object or attribute in the news stories (e.g., Kioussis, 2005; Littau & Stewart, 2015), and just a few have taken into account how redundant that object/attribute might be in terms of actual mentions (Moon, 2011). Consider a set of 10 news stories citing a certain attribute. If the attribute is cited only once in each story, the redundancy score is 10. But if it appears multiple times in all or most of the stories, the redundancy score is much higher and it can range widely among stories. If more than one attribute is being measured, the variance across attributes will be much broader as well. As such, redundancy may be a better measure of salience compared to the way frequency is often operationalized as presence/absence of an object or attribute.

The idea of redundancy in agenda-setting studies is not new. As McCombs (2014) has shown, a high degree of redundancy exists in the media agendas of even diverse news outlets. In this paper, I conduct two studies to explore if redundancy plays a different role in setting the public agenda when considering presence/absence of attributes in the stories, in comparison to the absolute mentions an attribute could get in the news coverage.

Finally, I use another measure, *degree centrality*, to analyze the effect of compelling arguments on public opinion. Drawing upon the Network Agenda Setting (NAS) model (Guo & McCombs, 2016), Saldaña and Ardèvol-Abreu (2016) suggest that degree centrality serves as another measure, in addition to frequency or redundancy, to understand why some attributes become compelling and others do not. Rather than exploring how frequently an item is covered, the NAS model (also known as the third level of the agenda-setting theory) focuses on how central an issue is located in relation to other issues on the media agenda. The centrality measure used in Guo, Chen, Vu, Wang, Aksamit, Guzek, Jachimowski, and McCombs' (2014) and Guo and McCombs' (2016) studies refers to the number of ties or links that a node has in a network. In other words, the more connected an element is with other elements in the news coverage, the more central the element is.¹

Degree centrality and frequency can be seen as related measurements. Yet, at least theoretically, an object or attribute's centrality does not necessarily correspond to its frequency in the media agenda (Guo, 2012). The most frequently reported issue may be mostly covered as a single issue in the news, thus the issue is not necessarily highly associated with other issues. Guo et al. (2014) and Saldaña

1. Although this article focuses on degree centrality, the NAS model also considers other network analysis measures, such as betweenness centrality, clusters and network density (Guo 2012).

and Ardèvol-Abreu (2016) have used centrality as an alternative measure of salience, finding intriguing and consistent results: it is not only the number of times an attribute is mentioned in the news, but also the association among different attributes, that help increase the public salience of an issue. Thus, in this paper, *frequency* and *degree centrality* are compared as two measures of salience to determine compelling arguments.

Big data tools for understanding media content

Computational methods offer the potential for overcoming some of the sampling and coding limitations of traditional content analysis (Lewis, Zamith & Hermida, 2013). With regard to coding, data can be examined using textual analysis and concept mapping tools that identify the most frequently used keywords and visualize their co-occurrence. Topic modeling is a good example of computerized content analysis. Based on the algorithm of Latent Dirichlet Allocation (LDA), the modeling process assumes there are hidden topics in textual bodies, and each topic is formed by some specific words that occur together more frequently than one would expect by chance (DiMaggio, Nag, & Blei, 2013). With this technique, researchers can find topics in large collections of text with more precision than other computerized tools. For example, Guo, Vargo, Pan, Ding, and Ishwar (2016) analyzed 77 million tweets about the 2012 U.S. presidential election. By comparing topic modeling with dictionary-based analysis, they found that LDA models were able to interpret more tweets and reveal more nuanced details of the conversation than the dictionary-based approach.

The use of computational tools is not without critiques. Many scholars have found them to yield satisfactory results only for surface-level analyses – manifest content –, sacrificing more nuanced meanings present in the analyzed texts – latent content – (Lewis et al., 2013). Manifest content is usually observable and countable, while latent content refers to constructs that are not directly observable or easily identifiable (Neuendorf, 2002). The manifest-latent dichotomy becomes evident when comparing computer-aided techniques with human judgment. Thus, the key to address these limitations is the type of content that is analyzed, and the goals of the study. Sjøvaag and Stavelin (2012) indicate that computer-assisted methods for the coding of manifest content has shown satisfactory results, acknowledging that “human labor is still considered superior for the coding of latent content” (p. 219). Similarly, Lewis et al. (2013) suggest that the *structural features* of media content can be more fully subjected to algorithmic analysis, while the *sociocultural contexts* built up around those features need the careful attention of manual methods.

A more practical (but not less important) limitation is the lack of training among social scientists to develop algorithms or scripts for computer-aided text analysis (Hasell & McGregor, 2016). In most cases, social researchers are required to do some extra work to adapt the tools developed by computer scientists (DiMaggio, 2015), which implies time and efforts to learn programs such as Python or R. In this paper, I propose a simple way to overcome this limitation with no need of advanced training in programming software. For the study of compelling arguments, I suggest the use of NVivo (a well-known text-analysis software that requires no coding training) to analyze key words that can be treated as attributes.

Previous studies have used words to represent objects and/or attributes. Bennett, Lawrence, and Livingston (2006) examined how the U.S. media portrayed the Abu Ghraib prison story in 2004 by looking at four words: mistreatment, scandal, abuse, and torture. The study involved machine coding of LexisNexis search results to ascertain the frequency of the words in the text corpus. Similarly, Kiouisis (2005) created a list of keywords to identify positive and negative aspects of presidential candidates in news media coverage, while Laver, Benoit and Garry (2003) used key words to extract policy positions from political texts.

Vargo, Guo, McCombs, and Shaw (2014) used words and computational analysis to identify the most salient issues in tweets about the 2012 U.S. presidential election. To identify the issues, they first identified the most common words in the entire corpus of tweets. By generating a frequency list, the authors manually examined all words that occurred more than a 1,000 times. Those words that corresponded directly to issues were then placed into issue construct lists: economy, foreign policy, individual liberties, federal programs, immigration, education, environment, and big government. The full corpus of tweets was content analyzed with computer-assisted tools, and then compared with the results of a manual content analysis. The results from this triangulation were found to be valid.

Content analysis requires a coding scheme that in turns implies researchers already know what is worth finding in the texts, before having analyzed those (DiMaggio et al., 2013). Vargo et al. (2014) overcame this difficulty by using words contained in their data – the issues they analyzed emerged from the data instead of a different source. In this paper, I follow Vargo et al.'s method and use words to identify the most important attributes of a particular issue. Similarly, I compare manual with computer-aided content analysis to assess the validity of both techniques.

Hypothesis and research questions

This study applies both manual content analysis (Study 1) and computational techniques (Study 2) to analyze the compelling argument concept, correlating both frequency and degree centrality measures with public opinion. To test this approach, I focus on the issue *unemployment*, which is a frequent issue on the public agenda (McCombs & Zhu, 1995) and it showed increasing prominence in the Gallup Polls during 2011. Unemployment is in itself interesting, as research has shown it can be either obtrusive or unobtrusive, depending on people's personal experiences (McCombs, 2014). Rather than being a stable issue in the public opinion, unemployment showed periods of valleys and peaks during 2011, as illustrated in Figure 1. Then, I observed the issue unemployment to identify if these variations could be related to media coverage.

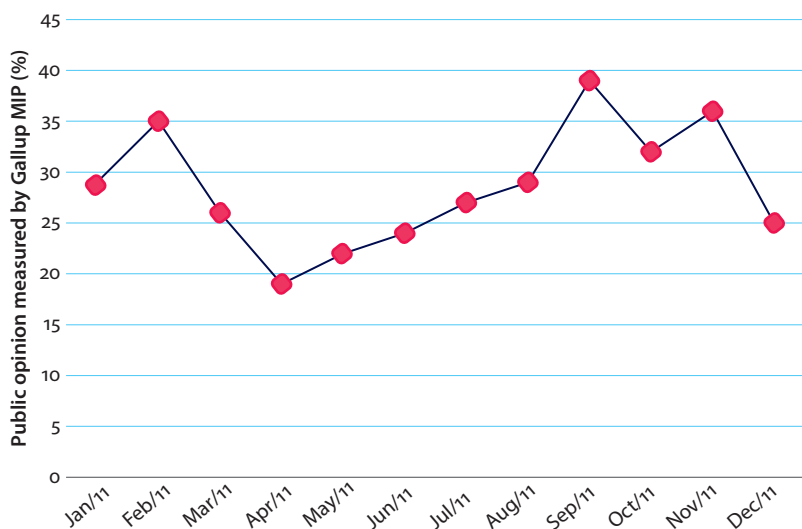


Figure 1. Public opinion about unemployment (monthly)

In the tradition of previous agenda-setting studies, the first hypothesis tests the relationship between news coverage of unemployment and the public's evaluation of that issue's importance.² Therefore,

H1: Media coverage of unemployment will be positively correlated with public opinion about that issue.

2. According to the Acapulco Typology for agenda-setting studies, this article belongs to the Type III category, as it focuses on a single issue with aggregate measures of the public agenda (McCombs, Holbert, Kioussis & Wanta, 2011).

In this paper, I explore if frequency – measured as presence/absence of an attribute in a story – plays a different role in setting the public agenda in comparison to the absolute mentions an attribute could get in the news coverage. Thus:

*RQ1a: Which attributes of the unemployment issue in the media agenda served as compelling arguments – if determined by **frequency** – for issue salience on the public agenda?*

*RQ1b: Which attributes of the unemployment issue in the media agenda served as compelling arguments – if determined by **absolute frequency** – for issue salience on the public agenda?*

As stated above, previous studies have incorporated degree centrality, in addition to frequency, as a tool to identify compelling arguments. To compare both frequency and degree centrality as valid measures to test compelling arguments, I ask:

*RQ2a: Which attributes of the unemployment issue in the media agenda served as compelling arguments – if determined by **degree centrality** – for issue salience on the public agenda?*

*RQ2b: Which attributes of the unemployment issue in the media agenda served as compelling arguments – if determined by **estimated degree centrality** – for issue salience on the public agenda?*

*RQ3a: What are the differences in how **frequency** and **degree centrality** predict compelling arguments?*

*RQ3b: What are the differences in how **absolute frequency** and **estimated degree centrality** predict compelling arguments?*

I conducted two studies to explore the advantages of computerized content analysis over manual content analysis in attribute agenda setting research. In Study 1, I applied manual content analysis to answer research questions 1a, 2a and 3a, while I used computer-assisted analysis in Study 2 to answer research questions 1b, 2b and 3b.

A specific method was developed for each study and it is described below.

Study 1

Method

This study identifies unemployment-related attributes, as well as the combination of certain attributes, that led to the growing salience of unemployment as an issue in the mind of the public during 2011. As such, the dependent variable in this study is the salience of the issue in public opinion. I used public opinion data collected by

the Gallup Most Important Problem (MIP) Polls, which provided the percentage of respondents who mentioned “unemployment” as the most important problem facing the country between January and December 2011.

To measure news coverage of the issue unemployment in 2011, I content analyzed news stories related to unemployment published on the front pages of *The New York Times*. A large number of agenda-setting studies have used this newspaper and its front page coverage to represent the U.S. media agenda. Known as an “elite newspaper of record” (Riffe et al., 1994, p. 16), *The New York Times* sets the agenda for many other media (McCombs 2004, 2014).

All the newspaper’s front-page stories about unemployment published in 2011 were retrieved from the LexisNexis Academic database. According to McCombs (2014), the empirical evidence suggests that the public agenda reflects the media agenda of the previous one or two months. Wanta and Hu (1994) found that the optimal agenda-setting time-lag for newspapers ranged from three to four weeks. Therefore, I searched the month prior to each poll for all stories that mentioned the word “unemployment” applying two filters to retrieve articles. First, I filtered for stories published in the *Times*’ front page, i.e., “Section A; Pg. 1,” and second, I filtered for stories published in both “National desk” and “Business/Financial desk,” leaving out stories related to international issues (“Foreign desk”). Following this criteria, 155 news stories were retrieved. Table 1 illustrates the distribution of stories by month.

Table 1. Unemployment-related stories in *The New York Times* in 2011. Distribution by month

January	13	May	10	September	20
February	10	June	10	October	18
March	11	July	11	November	18
April	7	August	12	December	15

As stated before, Vargo et al.’s (2014) method was adapted to create a list of attributes for the issue unemployment. I used NVivo, a text analysis software, to run a word frequency query and find the most common terms in the 155 stories. The initial list of terms included 280 words with a frequency equal or higher than 100 among the corpus of stories. However, most of these words were not directly related to unemployment. A qualitative, manual process was applied to select unemployment-related words only, such as “job,” “work,” or “company.” In addition, words not directly related to unemployment such as “Obama,” “Democrats,” or “Conservatives” were also selected, as they provide context for the issue. Although the media are the main agenda setters for objects and attributes on the public agenda, political actors also play a role in defining what issues are important for public

opinion (Son & Weaver, 2005) and the presence/absence of these actors might increase the salience of the issue in the public's mind. After this selection process, the initial list of most frequent terms was reduced to 73 words (see Table 2).

Table 2. Frequency of the most common unemployment-related words

Jobs	828	Debts	323	Leaders	215	Employees	133
Republicans	750	Plans	286	Politics	214	Consuming	131
Taxing	608	Benefits	280	Support	209	Funds	128
Obama	595	Programs	279	Recession	207	Security	128
Company	560	Pays	272	Washington	206	Family	126
Unemployment	521	Growth	261	Costs	205	Rising	126
Government	514	Senate	256	Pricing	204	Payrolls	124
Cutting	492	Markets	255	Bills	200	Stocks	118
Work	476	Economists	250	Hiring	198	Crisis	116
Housing	467	Money	247	Raise	193	Conservatives	112
President	467	Congress	238	Income	192	Investments	109
Economy	459	Employment	238	Whites	175	Declined	104
Economic	422	Banks	234	Romney	167	Living	104
Federal	415	Policy	232	Labor	166	Revenues	104
Spends	414	Help	229	Unions	162	Aid	103
Workers	360	Budget	227	Deficit	149	Credits	102
Democrats	350	Party	225	Reduce	136	Efforts	101
Business	336	Financial	220	Industry	134	Borrow	100
Increase	335						

The list of words in Table 2 was treated as a list of attributes describing (or giving context to) the issue unemployment in the news coverage. To identify compelling arguments in the news stories, a new NVivo word frequency query was run for each story, searching for the 73 attributes previously described. If the attribute was found in the story, it was coded as 1. If the attribute was absent, it was coded as 0. Then, I added the number of stories per month mentioning the attribute. If a certain month had seven stories about unemployment, and a certain attribute was found in five stories, then that month was coded as 5. This allowed to create a database with 12 months and 73 variables where the cell entries corresponded to the attribute's frequency in each month.

Data analysis

When testing the first-level agenda-setting hypothesis, the independent variable is the media coverage of the issue unemployment during 2011, measured by *the number of news stories about unemployment* published during the month prior

to each Gallup MIP survey. A zero-order Pearson's correlation was performed to explore the relationship between the media and the public agenda regarding unemployment.

When testing second-level agenda-setting effects, the independent variable of this study is the salience of the unemployment-related attributes in the news coverage, measured by *the number of news stories that mentioned a specific attribute* during the month prior to each survey. To identify the attributes that served as compelling arguments (RQ1a), zero-order Pearson's correlations were performed to examine the relationship between the frequency of the attributes in the news stories and the salience of the issue in the public agenda.

RQ2a asks what attributes on the media agenda served as compelling arguments when considering degree centrality. In this case, the independent variable is the salience of the attributes measured by its degree centrality, i.e., *the number of connections a specific attribute had with other attributes* during the month prior to each survey. Previous studies have used Ucinet, a network analysis software, to calculate all the possible connections an attribute may have in the same story (Guo & McCombs, 2011a, 2011b; Guo et al., 2014; Saldaña & Ardèvol-Abreu (2016). Therefore, I used Ucinet to calculate the degree centrality of each attribute. Each attribute's degree centrality score was correlated with the salience of the issue in the public agenda.

To answer RQ3a, I compared the results from RQ1a and RQ2a.

Results

The first hypothesis predicts that an increase in the media coverage of unemployment will be associated with an increase in the proportion of survey respondents naming this issue as the country's most important problem. Results show a significant, positive correlation between the number of stories per month and the Gallup MIP data ($r = .759, p < .01$). As expected, this hypothesis was supported – the more the media covered unemployment, the more the public saw it as an important issue.

RQ1a asked what attributes of the issue unemployment served as compelling arguments if determined by frequency of the attribute. Results show that 40 out of the 73 attributes served as compelling arguments – the more often these 40 attributes appeared in the stories, the more people mentioned unemployment as the most important problem facing the country. Table 3 (column 1) illustrates zero-order Pearson's correlations between the frequency of the attributes in the media agenda and the salience of the issue in the public agenda. The correlation coefficients for the 40 significant attributes are strong, ranging from .58 to .90 (median = .70).

RQ2a asked what attributes served as compelling arguments if determined by degree centrality. Results show that 48 out of the 73 attributes became compelling – the more central these attributes were in the attribute network, the more the issue unemployment resonated in people's mind. Table 3 (column 2) illustrates zero-order Pearson's correlations between the degree centrality of the compelling attributes in the media agenda, and the salience of the issue in the public agenda. The correlation coefficients for the 48 significant attributes are again strong, ranging from .58 to .89 (median = .70).

Table 3. Zero-order Pearson's correlations for Study 1.
Compelling arguments and public opinion (PO)

Attributes	Frequency & PO	Degree centrality & PO
Support	.904**	.892**
Bills	.883**	.883**
Leaders	.824**	.844**
Republicans	.805**	.821**
Economy	.787**	.814**
Company	.772**	.798**
Washington	.766**	.743**
Policy	.761**	.781**
Crisis	.742**	.755**
Federal	.740**	.753**
Increase	.734**	.781**
Unemployment	.733**	.776**
Deficit	.725**	.745**
Democrats	.719**	.796**
Aid	.706*	.727**
Congress	.703*	.699*
Politics	.703*	.749**
Economic	.702*	.746**
Help	.701*	.755**
Government	.700*	.720**
Obama	.696*	.686*
Housing	.695*	.697*
Conservatives	.692*	.628*
Plans	.690*	.684*
Raise	.688*	.735**
Financial	.681*	.747**
Markets	.652*	.672*
Business	.645*	.684*
Workers	.642*	.716**
Recession	.641*	.742**

(continued)

Table 3. (continued)

Attributes	Frequency & PO	Degree centrality & PO
Debts	.636*	.655*
Banks	.617*	.588*
Jobs	.615*	.660*
Programs	.603*	.598*
Cutting	.595*	.668*
Stocks	.595*	.618*
Growth	.591*	.660*
Work	.588*	.653*
Spends	.580*	.637*
Romney	.595*	.508
Tax	0.565	.656*
Labor	0.553	.642*
Budget	.509	.636*
Funds	.566	.615*
Industry	.504	.609*
Whites	.523	.600*
Economists	.518	.589*
President	.571	.586*
Reduce	.518	.582*

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

RQ3a asked whether frequency and degree centrality would predict compelling arguments differently. According to Table 3, frequency identified 40 attributes as compelling arguments, 39 of which were also identified by degree centrality.³ When correlating degree centrality, another nine attributes became compelling – these attributes behaved as compelling arguments only when mentioned in tandem with other attributes.

When looking at correlation coefficients, the strongest correlations with public opinion are Support ($r = .904, p < .01$), Bills ($r = .883, p < .01$), Leaders ($r = .824, p < .01$), and Republicans ($r = .805, p < .01$) if correlating frequency. Interestingly, these attributes are also at the top of the list when correlating degree centrality (Support: $r = .892, p < .01$; Bills: $r = .883, p < .01$; Leaders: $r = .844, p < .01$; Republicans: $r = .821, p < .01$). These findings show a strong consistency between both measures.

3. “Romney” was identified as compelling by frequency but not by degree centrality (see Table 3, columns 1 and 2).

Study 2

Method

Study 2 aims to further extend the compelling-argument hypothesis to the “big data” landscape. By using computerized content-analysis and mathematical techniques, this study’s goal is to develop an alternative approach to content-analyze news stories and to obtain a more precise measure of the attribute’s redundancy over time.

Data analysis

To make a valid comparison between Study 1 and Study 2, I analyzed the same 155 news stories, yet applying different methods. In Study 1, each story was content-analyzed individually, using a word frequency query to find specific attributes in that story. In Study 2, I grouped stories by month, in order to run a frequency query for each month instead of one for each story. As seen in Table 1, April had the lowest number of news stories (7) while September had the highest (20). With this method, however, the number of stories per month could be much larger and still only one frequency query would be needed. Therefore, this study is not a “big data” study itself, as the number of analyzed stories is very small. Rather, this study aims to provide a method for a fast analysis of large text corpuses without the use of programming software.

Secondly, Study 1 coded for the presence/absence of an attribute in each story, regardless how many times the attribute was actually mentioned in the story – the word was either present or absent. In Study 2, I did consider this difference and coded for *absolute frequency* – the exact number of times the word was mentioned each month. In consequence, the word frequency query searched for the absolute number of times the attribute appeared in the stories that month. To answer RQ1b, absolute frequency was correlated with the salience of unemployment in the public agenda.

Third, I coded for degree centrality using an alternative calculation. When working with aggregate data, as in the case of big data analysis, the information of small units of analysis is not always available. In Study 2, I did not analyze each story separately, but grouped all stories by month and used *the month* as a unit of analysis. Degree centrality, the measure used in Study 1, is a measure of connectedness – in this case, how connected an attribute is in each story. Since the unit of analysis is the month instead of the story, it is not possible to calculate the attributes’ connections in each story. Therefore, degree centrality calculation, as it has been calculated in previous studies, is not feasible.

To overcome these limitations, an alternative measure of degree centrality was calculated. First, I calculated the average number of attribute words per month⁴

4. Considering the words that were treated as attributes only.

and divided this by the number of stories in each month. Then, I adjusted the frequency of each word by the average of words per story, per month. This is the estimation of the number of stories where the word might have appeared. After that, I calculated the probability of two attributes being together in a story, in order to find the probability number of the total connections an attribute might have. For instance, when comparing *Jobs* with *Cutting* in January 2011, probability calculation showed *Jobs* was distributed in the 13 stories found in January,⁵ while *Cutting* was present in only five. Then, I calculated the probability that *Jobs* and *Cutting* appeared together in five stories or less by using the following formula:

$$P(k) = \frac{\binom{n_1}{k} \binom{N - n_1}{n_2 - k}}{\binom{N}{n_2}}$$

Where:

N = number of stories in that month

n_1 = number of stories mentioning attribute 1⁶

n_2 = number of stories mentioning attribute 2

k = number of times attributes 1 and 2 are mentioned together

At this point, the calculation of the expected value of an attribute's connections is straightforward. The formula to add all numbers of connections by the probability of occurrence of each connection is:

$$E(k) = \sum_{k=1}^{\min(n_1, n_2)} k \cdot P(k)$$

The above formula predicted the number of connections two attributes might have. This process was repeated for the 73 attributes in the 12 months, to create a database of the degree centrality estimate of each attribute. Then I correlated the estimated degree centrality with the salience of the issue in the public agenda to answer RQ2b.

As in Study 1, the results from RQ1b and RQ2b were compared to answer RQ3b.

5. See Table 1 to find the number of stories per month.

6. When comparing two attributes, Attribute 1 must be the one with the highest absolute frequency.

Results

RQ1b asked what attributes of the issue unemployment become compelling if determined by absolute frequency.⁷ Results show that 11 out of the 73 attributes served as compelling arguments. In other words, the more these 11 attributes appeared in the media agenda, the more unemployment became an important issue in the public agenda. The correlation coefficients for the 11 significant attributes ranged from .58 to .76 (median = .67) (see Table 4, column 1).

Regarding RQ2b about what attributes are compelling if determined by estimated degree centrality, results show that 28 out of the 73 attributes became compelling. The correlation coefficients for the 28 significant attributes ranged from .60 to .74 (median = .73) (see Table 4, column 2).

Table 4. Zero-order Pearson's correlations for Study 2.
Compelling arguments and public opinion (PO)

Attributes	Absolute frequency & PO	Estimated degree centrality & PO
Washington	.762**	.735**
Economy	.744**	.729**
Bills	.720**	.725**
Policy	.680*	.659*
Politics	.680*	.659*
Republicans	.666*	.729**
Plans	.644*	.714**
Crisis	.640*	.580*
Leaders	.640*	.624*
Banks	.627*	.618*
Congress	.584*	.608*
Taxing	.300	.756**
President	.538	.737**
Obama	.471	.737**
Jobs	.419	.729**
Work	.346	.729**
Economic	.535	.729**
Democrats	.423	.726**
Government	.518	.721**
Spends	.531	.717**
Housing	.424	.713**
Unemployment	.533	.703*

7. I didn't test H1 again because the variables were the same in Study 1 and Study 2: number of news stories and Gallup MIP survey.

Table 4. (continued)

Attributes	Absolute frequency & PO	Estimated degree centrality & PO
Federal	.515	.697*
Money	.420	.679*
Debts	.261	.665*
Company	-.024	.657*
Increase	.500	.638*
Party	.512	.604*

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

RQ3b asked whether absolute frequency and estimated degree centrality predict compelling arguments differently. According to Table 4 (columns 1 and 2), absolute frequency predicted 11 attributes as compelling arguments, which were also predicted by estimated degree centrality: Washington, Economy, Bills, Policy, Politics, Republicans, Plans, Crisis, Leaders, Banks, and Congress. When correlating estimated degree centrality, another 18 attributes became compelling: Taxing, President, Obama, Jobs, Work, Economic, Democrats, Government, Spends, Housing, Unemployment, Federal, Money, Debts, Company, Increase, and Party. These attributes behave as compelling only when mentioned together with other attributes.

When looking at the correlation coefficients, the strongest correlations with public opinion are Washington ($r = .762, p < .01$), Economy ($r = .744, p < .01$), and Bills ($r = .720, p < .01$) when correlating absolute frequency. However, when correlating estimated degree centrality, these attributes are not at the top of the list, and instead Taxing ($r = .756, p < .01$), President ($r = .737, p < .01$) and Obama ($r = .737, p < .01$) surfaced. These attributes were not compelling when considering absolute frequency in the correlation.

Discussion

Theoretical connotations: Absolute frequency to understand salience

Study 1 found 40 attributes serving as compelling arguments if correlating presence/absence of attributes in the stories. If correlating degree centrality, 48 attributes became compelling, which is about two thirds of the total number of attributes. The reason so many compelling arguments were found relies on the way the attributes of the study were selected. One of the criticisms toward content analysis is the assumption that the researcher knows *a priori* what variables to look at, mostly based on previous research. Thus, to avoid the creation of arbitrary categories, I built a list of attributes that emerged from the actual data. Since these

attributes came from a list of the most frequent words in the corpus of news stories, the probability of these words (attributes) not being in the stories is very low. When coding for frequency in terms of presence/absence, most of the words were present at least once in each story. In consequence, there was not a great variance and many attributes behaved in the same manner. The same is true when considering degree centrality.

Study 2 applied a different measure of redundancy – the absolute frequency an attribute was mentioned each month. If we accept the hypothesis that more media attention leads to more public attention toward a certain issue, then we cannot ignore redundancy as a measure of media attention. Just a few studies have counted the absolute frequency of attributes (Moon, 2011, for instance), but to my knowledge, this is the first attempt to test whether repeating a specific word (attribute) in the same story might make a difference. This study shows that coding only for presence or absence increases the risk of missing the nuances related to redundancy. Similarly, it also increases the probability of making a type I error. Coding for presence/absence gives all attributes the same status – they are present in the stories, or they are not. It does not consider how many times an attribute is mentioned, and therefore, it does not measure the real salience an attribute may have. In consequence, the correlations show relationships that vanish when a more accurate measure of salience is accounted for.

The list of compelling arguments in Study 1 is therefore so long, that it is quite hard to interpret. In Study 2, in contrast, we observe a more parsimonious list of compelling arguments, which in turns allows for a more meaningful interpretation. When observing the specific words that behaved as compelling arguments, we see that attributes referred to institutions or macro elements, such as *Economy*, *Politics*, *Crisis* or *Banks*, resonated more on people's minds. However, when analyzing the estimated centrality of these attributes, other words surfaced. At this level, attributes related to the U.S. government acted as compelling, such as *Obama*, *President*, *Government* or *Federal*. So, it seems that people do not directly relate the President's image with the unemployment problem, but when other attributes are mentioned, President Obama cannot be left out of the picture. Also, more "positive" words are significantly related to public opinion at this level, such as *Jobs*, *Works*, *Spending* or *Increase*. Therefore, it is possible that, in people's minds, the President is closer to the 'solutions' for unemployment, than to the 'causes' of the problem. If that is the case, these findings confirm previous studies that have found that negative news are more successful than positive news in catching people's attention, and that is why "unemployment problems" showed up before "unemployment solutions."

In general terms, results showed a high, significant correlation between news coverage and public opinion at the most basic level of the agenda-setting theory. Confirming what hundreds of studies have previously found, there is a transfer

of salience from the media agenda to the public agenda regarding a specific object – in this case, unemployment. Likewise, the analysis of the attribute agenda, represented by the most redundant unemployment-related words, confirmed the idea that certain attributes resonate more than others – compelling arguments.

Methodological contributions: A new approach for compelling arguments

By comparing human and computer-assisted content analysis, this study makes novel methodological claims. First, the application of computerized techniques to identify attributes proved to be efficient, since it was able to find words describing the issue (i.e., object's attributes) and these words showed a high, significant relationship with public opinion. Then, the frequency of these words changed in the same direction that people's opinion did, validating once again the hypothesis that people learn from the media, and the more the media repeat a message, the more people pay attention and give importance to that message. The word frequency query through NVivo was able to successfully capture this relationship.

Second, the estimation of degree centrality in Study 2 provided an alternative approach to measure attributes' network relationships. Results from the correlations using a centrality estimate were consistent with the correlations using absolute frequency. In other words, results made sense. This is a great achievement, since I did not consider the number of times a word (attribute) was mentioned in a specific story, or how many stories mentioned two attributes together. What I had was a set of stories per month, and the number of times an attribute was mentioned that month. Then, the estimation referred to *the probability* of an attribute being mentioned in n stories, and *the probability* of two attributes co-occurring in one story (or more). The use of probabilities provided a degree centrality estimate for each attribute, which worked in the direction hypothesized.

This does not imply, by any means, that content analysis of news stories does not require human coders any longer. First, NVivo provided a list of the most frequent words with a great number of non-unemployment-related terms. Human coding was needed to select the final list of words (attributes) to move to the next step of the analysis.

Practical implications: Analysis of big data

When having a manageable set of stories, more complex attributes can be created, and frequency (understood as presence/absence) may work well to predict compelling arguments. But in studies examining large data sets, where stories number in the thousands or millions, a word list is a more parsimonious solution to find attributes, and that is why absolute frequency works better than traditional frequency. As stated

by Lewis et al. (2013), computer content analysis is a good technique for the coding of manifest content, as it is the case in this paper, bringing a tighter focus on those attributes that function as compelling arguments. For big data projects focused on identifying structural features of media content, the method presented in this study provides a simple way to content analyze large collections of text.

Limitations and future research

This study focused on the issue unemployment, which can be obtrusive for some individuals but unobtrusive for others: those without a job do not need the media to tell them unemployment is relevant, as the issue is obtrusive for them. Future research should test the innovations proposed on this study on issues with different levels of obtrusiveness (e.g., immigration, climate change, foreign policy) to analyze how different sets of words can become compelling arguments. Similarly, future papers should look at clusters of words, in order to see if attributes similar to each other might have a higher likelihood of becoming compelling arguments.

Acknowledgments

The author thanks Dr. Max McCombs for his generous guidance and support while writing this article, and Paro Pain for her valuable help with the manual content analysis.

References

- Bennett, W. L., Lawrence, R. G., & Livingston, S. (2006). None dare call it torture: Indexing and the limits of press independence in the Abu Ghraib scandal. *Journal of Communication*, 56(3), 467–485. doi:10.1111/j.1460-2466.2006.00296.x
- Berelson, B. (1952). *Content analysis in communication research*. New York, NY: Free Press.
- Bowe, B.J., Fahmy, S., & Wanta, W. (2013). Missing religion: Second level agenda setting and Islam in American newspapers. *The International Communication Gazette*, 75(7), 636–652. doi:10.1177/1748048513482544
- Chyi, H. I., & McCombs, M. (2004). Media salience and the process of framing: Coverage of the Columbine school shootings. *Journalism and Mass Communication Quarterly*, 81(1), 22–35. doi:10.1177/107769900408100103
- DiMaggio, P., Nag, M. & Blei, D. (2013) Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*, 41, 570–606. doi:10.1016/j.poetic.2013.08.004
- DiMaggio, P. (2015). Adapting computational text analysis to social science (and vice versa). *Big Data & Society*. doi:10.1177/2053951715602908
- Ghanem, S. I. (1996). *Media coverage of crime and public opinion: An exploration of the second level of agenda setting*. Unpublished doctoral dissertation, University of Texas at Austin.

- Ghanem, S. I. (1997). Filling in the tapestry: The second level of agenda setting. In M. E. McCombs, D. Shaw & D. Weaver (Eds.), *Communication and democracy* (pp.309–330). Mahwah, NJ: Lawrence Erlbaum.
- Guo, L. (2012). The application of social network analysis in agenda setting research: A methodological exploration. *Journal of Broadcasting & Electronic Media*, 56(4), 616–631. doi:10.1080/08838151.2012.732148
- Guo, L., Chen, Y., Vu, H., Wang, Q., Aksamit, R., Guzek, D., Jachimowski, M. & McCombs, M. (2014). Coverage of the Iraq War in the United States, Mainland China, Taiwan and Poland. *Journalism Studies*. doi:10.1080/1461670X.2014.885198
- Guo, L. & McCombs, M. (eds.) (2016). *The power of information networks: New directions for agenda setting*. New York: Routledge.
- Guo, L., & McCombs, M. (2011a). *Network agenda setting: A third level of media effects*. query Paper presented at the International Communication Association (ICA) Annual Conference, Boston.
- Guo, L., & McCombs, M. (2011b). *Toward the third level of agenda setting theory: A network agenda setting model*. Paper presented at the Association for Education in Journalism and Mass Communication Annual Conference, St. Louis, MO.
- Guo, L., Vargo, C. J., Pan, Z., Ding, W., & Ishwar, P. (2016) Big social data analytics in journalism and mass communication: Comparing dictionary-based text analysis and unsupervised topic modeling. *Journalism & Mass Communication Quarterly*, 1–28.
- Hasell, A. A., & McGregor, S. C. (2016). *Black-box algorithms: Scholarly use of proprietary coding software in communication research*. Unpublished manuscript.
- Kim, Y. & Kim, Y. (2014) *Theoretical and methodological trends of agenda setting theory: A thematic meta-analysis of the last five decades*. Working paper, University of Alabama.
- Kiousis, S. (2005). Compelling arguments and attitude strength. Exploring the impact of second-level agenda setting on public opinion of presidential candidate images. *The International Journal of Press/Politics*, 10(2), 3–27. doi:10.1177/1081180X05276095
- Laver, M., Benoit, K., & Garry, J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review*, 97(2), 311–331. doi:10.1017/S0003055403000698
- Lewis, S. C.; Zamith, R. & Hermida, A. (2013). Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of Broadcasting & Electronic Media*, 57(1), 34–52. doi:10.1080/08838151.2012.761702
- Littau, J. & Stewart, D. R. (2015). “Truthiness” and second-level agenda setting: Satire news and its influence on perceptions of television news credibility. *Electronic News*, 9(2), 122–136. doi:10.1177/1931243115581416
- McCombs, M. (2004). *Setting the agenda: The mass media and public opinion*. Cambridge, England: Polity Press.
- McCombs, M. (2014). *Setting the agenda: The mass media and public opinion* (2nd edition). Cambridge, England: Polity Press.
- McCombs, M. & Ghanem, S. (2001). The convergence of agenda setting and framing. In S. D. Reese, O. H. Gandy Jr. & A. E. Grant (Eds.), *Framing public life: Perspectives on media and our understanding of the world* pp. 67–82). Mahwah, NJ: Lawrence Erlbaum.
- McCombs, M., Holbert, L., Kiousis, S., & Wanta, W. (2011). *The news and public opinion: Media effects on civic life*. Cambridge: Polity Press.
- McCombs, M., Llamas, J. P., Lopez-Escobar, E., & Rey, F. (1997). Candidate images in Spanish elections: Second-level agenda-setting effects. *Journalism and Mass Communication Quarterly*, 74(4), 703–717. doi:10.1177/107769909707400404

- McCombs, M. & Zhu, J. (1995). Capacity, diversity and volatility of the public agenda: Trends from 1954 to 1994. *Public Opinion Quarterly*, 59(4), 495–525. doi:10.1086/269491
- McMillan, S. J. (2000). The microscope and the moving target: The challenge of applying content analysis to the World Wide Web. *Journalism & Mass Communication Quarterly*, 77(1), 80–98. doi:10.1177/107769900007700107
- Moon, S. J. (2011). Attention, attitude, and behavior: Second-level agenda-setting effects as a mediator of media use and political participation. *Communication Research*, 40(5), 698–719. doi:10.1177/0093650211423021
- Neuendorf, K. (2002). *The content analysis guidebook*. California: Sage.
- Riffe, D., Lacy, S. R., & Fico, F. G. (2005). *Analyzing media messages: Using quantitative content analysis in research* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Saldaña, M. & Ardèvol-Abreu, A. (2016). From Compelling Arguments to Compelling Associations at the Third Level of Agenda Setting. In L. Guo & M. McCombs (Eds.), *The Power of Information Networks: New Directions for Agenda Setting*. (pp. 104–118). New York: Routledge.
- Sheafer, T. (2007) How to evaluate It: The role of story-evaluative tone in agenda setting and priming. *Journal of Communication*, 57, 21–39.
- Sjøvaag, H., & Stavelin, E. (2012) Web media and the quantitative content analysis: Methodological challenges in measuring online news content. *Convergence: The International Journal of Research into New Media Technologies*, 18(2), 215–229.
- Son, Y. J. & Weaver, D. (2005). Another look at what moves public opinion: Media agenda setting and polls in the 2000 U.S. election. *International Journal of Public Opinion Research*, 18(2), 174–197. doi:10.1093/ijpor/edh090
- Vargo, C. J., Guo, L., McCombs, M., & Shaw, D. L. (2014). Network issue agendas on Twitter during the 2012 U.S. presidential election. *Journal of Communication*, 64, 296–316. doi:10.1111/jcom.12089
- Wanta, W. & Hu, Y. (1994). Time-lag differences in the agenda-setting process: An examination of five news media. *International Journal of Public Opinion Research*, 6(3), 225–240. doi:10.1093/ijpor/6.3.225
- Yioutas, J. & Segvic, I. (2003). Revisiting the Clinton/Lewinsky scandal: The convergence of agenda Setting and framing. *Journalism & Mass Communication Quarterly*, 80(3), 567–82. doi:10.1177/107769900308000306

Author's address

Magdalena Saldaña
 Texas Tech University
 College of Media & Communication
 3003 15th Street, Box 43082
 Lubbock, TX 79409-3082
 USA
 Magdalena.saldana@ttu.edu