Psychology and Marketing, Volume 34, Issue 3, 2017, pp. 275–293
DOI:10.1002/mar.20988

ANTECEDENTS OF RETWEETING IN A (POLITICAL) MARKETING CONTEXT

Lorna Walker, Cranfield University

Paul R. Baines, Cranfield University (Corresponding author)

Radu Dimitriu, Cranfield University

Emma K. Macdonald, Cranfield University

Running Head: Antecedents of Retweeting

ABSTRACT

Word of mouth disseminates across Twitter by means of retweeting; however the antecedents of retweeting have not received much attention. This study uses the CHAID decision tree predictive method (Kass, 1980) with readily available Twitter data, and manually coded sentiment and content data, to identify why some tweets are more likely to be retweeted than others in a (political) marketing context. The analysis includes four CHAID models: (i) using message structure variables only, (ii) source variables only, (iii) message content and sentiment variables only and (iv) a combined model using source, message structure, message content and sentiment variables. The aggregated predictive model correctly classified retweeting behavior with a 76.7% success rate. Retweeting tends to occur when the originator has a high number of Twitter followers and the sentiment of the tweet is negative, contradicting previous research (East, Hammond, & Wright, 2007; Wu, 2013) but concurring with others (Hennig-Thurau, Wiertz, & Feldhaus, 2014). Additionally, particular types of tweet content are associated with high levels of retweeting, in particular those tweets including fear appeals or expressing support for others, whilst others are associated with very low levels of retweeting, such as those mentioning the sender's personal life. Managerial implications and research directions are presented. The study makes a methodological contribution by illustrating how CHAID predictive modelling can be used for Twitter data analysis and a theoretical contribution by providing insights into why retweeting occurs in a (political) marketing context.

KEYWORDS: retweeting behavior; social media analysis; CHAID analysis; voter engagement; political marketing; Twitter; WOM

INTRODUCTION

Set in the context of a (political) marketing campaign, this study investigates why some Twitter messages are retweeted whilst others are not. Twitter, a form of 'micro-blogging', enables rapid message dissemination. Consequently, it has played a central role in recent high profile political events, including campaigns to generate civil unrest in Moldova (2009), promote Iranian election protests (2009-10) and advance the Tunisian and Egyptian revolutions of 2010-11(Hermida, Lewis, & Zamith, 2014; Lotan et al., 2011). In addition, much of President Obama's success in 2008 (the 'Yes we can!' campaign) was attributed to his ability to harness social media power including Twitter (Graham, Broersma, & Hazelhoff, 2013) and Twitter is now the social media network of choice for politicians and journalists. Newspapers and other offline media have normalized the use of Twitter as source material, thus amplifying its impact, and many politicians use social media as a way of building relationships with the media (Newman, 2010) as well as for communication with voters. In January 2016, a quarter of American adults said they looked to candidates' social media posts for US election information (Mitchell, Holcomb, & Weisel, 2016). Of the many social networks available, Twitter is the most suited to this type of research because it is inherently public and outward-facing (unlike, for example, Facebook or WhatsApp which are both private by default meaning researchers can only access a very limited sample of posts). Nevertheless, although tweets can be highly influential, little of the dynamics behind message sharing on this platform (i.e., retweeting) is currently understood.

Twitter has experienced rapid growth since its inception in 2006. It reports having 313 million active users (in June 2016) sending more than 500 million tweets per day (internetlivestats, 2016), of which 82% were sent via mobile (see Twitter, 2016[1]). A user's tweets are, by default, publicly viewable on their personal profile page and in the Twitter

[1] The company has yet to update its tweets per day figure indicating that in mid 2016 this had yet to reach 600 million.

feeds of followers, facilitating their mining for research purposes. Whilst some researchers underpin their Twitter studies by considering the blogging literature (e.g. Larsson & Hallvard, 2011), Twitter is more appropriately conceived as a form of electronic word of mouth (eWOM), sharing greater commonality with this mode of communication than with long-format blogging (Jansen & Zhang, 2009).

Twitter's principal feature, making it particularly powerful as a form of eWOM, is the retweet. When someone retweets, they pass on the original message to their own followers, enabling wider dissemination. This amplifies the original message and can extend its reach hundreds or even thousands of times.

Much extant research uses statistical modelling to predict retweeting (e.g. Rudat, Buder, & Hesse, 2014; Suh et al., 2010; Tang et al., 2015; Webberley, Allen, & Whitaker, 2016) but these researchers provide minimal discussion of why particular tweets are retweeted and fail to consider the implications for social media practitioners. Extant research focuses on using easily appended and analyzed variables including: structural elements of a tweet (e.g., the number of hashtags in the tweet) or certain tweet sender characteristics (e.g., number of sender's followers). Previous studies have given only limited consideration to tweet content as a retweeting antecedent. In parallel literature, another body of research uses the content analysis of tweets to ascertain what issues people are tweeting about, particularly in a political context (e.g. Golbeck, Grimes, & Rogers, 2010; Graham et al., 2013; Hemphill, Otterbacher, & Shapiro, 2013). Such analysis is also partial, as researchers do not generally include consideration of retweets or what stimulates them. This study bridges the gap between these two literatures by using CHAID analysis (chi-square automatic interaction detection) to build predictive models for message retweeting using four types of predictors combining message characteristics including a) tweet structural elements, b) content of

tweets, c) sentiment of tweets, and d) source characteristics (i.e., characteristics of the tweet sender).

The article makes a primary contribution by developing and illustrating CHAID as a replicable methodological approach for analyzing retweeting antecedents, usable in a variety of scenarios (e.g. commercial and political marketing contexts). The study makes a secondary contribution by extending the literature examining the antecedents of retweeting behavior (in the 2015 British General Election).

The remainder of the paper is organized as follows. The literature review discusses retweeting in a political marketing context, drawing on the substantive body of work scrutinizing the factors impacting on communication effectiveness, especially source and message characteristics. These characteristics are antecedents in generating retweets and are specifically investigated in the current research. The following section discusses the methodological approach adopted in the study to develop a predictive model of retweeting. This is followed by a consideration of the findings including an analysis of the effects of source characteristics and tweet message structural elements, sentiment and content on retweeting. The section specifically debates the approach taken to manually code and derive sentiment and content categories. A discussion follows, considering how the study contributes to the literature on message retweeting alongside the practical implications of the findings for political and commercial marketing contexts. The discussion also debates the limitations of the current study and illustrates opportunities for further research.

LITERATURE REVIEW

Online political campaigning has professionalized significantly, but particularly since the advent of social media, with emphasis placed on party-based campaigns directed from the center (Lee, 2014). Parallel to this, individual politicians use the internet and social media

platforms for voter engagement. Whilst social media is yet to be universally adopted by UK politicians, its use is rising and Twitter is politicians' social network of choice. There was a nearly six-fold increase in Twitter uptake between July 2009, when 63 (9.8%) UK MPs had Twitter accounts (Jackson & Lilleker, 2011), and the 2015 General Election, when 366 (56.6%) MPs were active on Twitter. In the US, political Twitter usage has also grown significantly. On Election Day in 2008, the Obama campaign tweeted once to an audience of around one million followers. By contrast, on Election Day in 2012, the Obama campaign disseminated more than 300 tweets to over 27 million followers (Helm, 2013). Virtually all members of the US House of Representatives are now active on Twitter.

Social media, and Twitter in particular, offers new pathways for brand owners, including politicians, to interact with their publics and vice versa. It has the potential to enable individual politicians to communicate directly with constituents and to disseminate their messages as individuals as opposed to relying on party 'campaign machines' to communicate or to depending on journalists' mediated messages. Social media provides a way of reaching people not easily accessed by traditional media such as terrestrial television (Cook, 2013), and of generating widespread political activism (Davies, 2014). Obama's 'Yes we can' campaign for the 2008 US Presidential election (see Tumulty, 2007) demonstrated that social media can galvanize voter support and generate significant campaign funding (Davies, 2014). Politicians can also use social media to reduce the psychological distance between themselves and their constituents (Vergeer, Hermans, & Sams, 2010).

While being on Twitter may bring exposure to reputational risks (Campanile, 2013), campaign strategists advise politicians to embrace the medium (e.g. Agranoff & Tabin, 2011; Cook, 2013; Davies, 2014; Enli & Skogerbø, 2013; Ross & Burger, 2014). In particular, because this allows them to secure a wider network from which to disseminate their message

via retweeting. However, there is little research examining the extent to which politicians' tweets are retweeted and what factors impact such retweeting.

Communication effects in commercial and political marketing

The influential transmission model of communication conceives a sender (i.e., source) conveying a message over a channel to a receiver, with environmental noise as interference (Shannon & Weaver, 1949). Existing research on communication effectiveness allocates a great deal of interest to how such factors influence attitudes and persuasion (e.g., Karmarkar & Tormala, 2009; Kirmani & Shiv, 1998; Petty & Cacioppo, 1984), product purchase intentions (Gilly et al., 1998; Yan, Ogle, & Hyllegard, 2010) or actual sales (Chevalier & Mayzlin, 2006; Zhu & Zhang, 2010). The research on sharing marketing messages is even more scant, with extant studies considering message recommendation intentions (Lee & Youn, 2009), sharing online information (Chiu et al., 2014) or forwarding e-mails (Chiu et al., 2007). The current research looking at retweeting falls into this category.

Even though the communication channel (e.g., Kaid, 2002; Lee & Youn, 2009) and the communication environment (Chiu et al., 2007) have been shown to have important effects on communication effectiveness, research looking at the drivers of retweeting largely deals with other aspects. A first category of studies deals with receiver/individual aspects (e.g., boyd, Golder & Lotan, 2010; Kim, Sung & Kang, 2014; Rudat., Buder & Hesse, 2014; Yang et al., 2010). This resonates with other studies on how receiver characteristics, e.g., involvement (Petty & Cacioppo, 1984; Petty, Cacioppo, & Schumann, 1983), knowledge (Park & Kim, 2008) or cultural background (Park & Lee, 2009), impact communication outcomes.

A second category looks at message/tweet or source/tweet sender aspects (e.g. Bakshy et al., 2011; Kupavskii et al., 2012; Lemahieu et al., 2015; Petrovic, Osborne & Lavrenko,

2011; Suh et al., 2010). The current research falls in this category, as it endeavors to determine what message and source characteristics drive retweeting. Consequently, the sections below discuss commercial and political marketing research on source and message characteristics impacting communication effectiveness, including the retweeting of Twitter messages.

Source characteristics

A significant body of research looks at source characteristics impacting communication effectiveness. Most researchers agree two characteristics affect the credibility of an information source, expertise and trustworthiness (Berlo, Lemert, & Mertz, 1969; Hovland & Weiss, 1951; Ohanian, 1990). Source expertise refers to the perceived competence of the source providing the information; experts are perceived to be a source of valid assertions (Hovland & Weiss, 1951; Ohanian, 1990). Further, source trustworthiness refers to the possible bias/incentives reflected in the source's information (Eagly & Chaiken, 1993). A trustworthy person generates others' confidence by communicating the assertions they personally consider most valid (Ohanian, 1990). In general, people tend to evaluate the credibility of a communication source by looking at source identity-descriptive information (MacDonald & Elahee, 2003).Source credibility is also influential for retweeting. One important cue in this context is the number of one's Twitter followers, which people use as a heuristic to assess expertise and trustworthiness. Extant research consistently finds that the number of followers a sender has predicts how likely their tweets are to be retweeted (e.g. Suh et al., 2010). Jin and Phua (2014) suggest that when a celebrity has many followers, people see them *ceteris paribus* as a more reliable and attractive message source than those with fewer followers.

Nevertheless, the number of followers someone has on Twitter is a blunt influence measure. Based on a dataset of 1.7 billion tweets, from over 6 million Twitter users, Cha et al. (2010) compared the number of followers, retweets and mentions for each user and found that the most followed users were not necessarily the most influential in terms of generating retweets. It seems likely that other source characteristics beyond the number of followers play a role in generating retweets. Such a factor and indicator of source credibility can be whether or not one's account is verified. Twitter applies verified status to the accounts of well-known people once confident of their bona fides. Petrovic et al. (2011) found that 91% of verified users' tweets were retweeted compared to only 6% of non-verified users', a 15:1 likelihood ratio.

Beyond credibility, other source traits, e.g. source likability and celebrity status, influence communication effectiveness (e.g., Kahle & Homer, 1985; Kamins, 1990; Petty & Cacioppo, 1984). In a political marketing context, politicians often act as opinion leaders (i.e., especially with core audiences). Opinion leadership is defined as "the degree to which an individual is able informally to influence other individuals' attitudes or overt behavior in a desired way with relative frequency" (Rogers, 2003:300). The principle of social proof means people determine what is 'correct' by observing what others deem is 'correct' (Cialdini, 1993), and what well-regarded opinion leaders think and say is influential.

Followers are likely to have a range of knowledge about politicians. On Twitter, there are many other factors indirectly related to a politician's profile that might influence a perceived credibility or likability. For instance, evidence from America and Scandinavia suggests candidates' positions in electoral races influences the way in which they use Twitter (Larsson & Kalsnes, 2014), and evidence elsewhere demonstrates that party affiliation influences politicians' Twitter behavior (Jackson & Lilleker, 2011). Consequently, it is sensible to consider whether or not factors including gender, age, party affiliation and

marginality[2] of seat also influence how followers respond to politicians' tweets. By including these factors, this study goes significantly beyond extant research on the role played by source characteristics in determining retweets, which have focused on information gleaned from Twitter only.

Message characteristics

Research on what message characteristics influence communication effectiveness considers both message content (Karmarkar & Tormala, 2009) and message positivity/negativity (Berger & Milkman, 2012; Chevalier & Mayzlin, 2006). For the case of Twitter messages, tweet characteristics can be broadly grouped into three categories: (i) the structural elements of the tweet (e.g., inclusion of hashtags, links, mentions of other Twitter users, whether a post or a reply, etc.), (ii) the sentiment of the tweet (positive, negative or neutral in tone) and (iii) the content of the tweet (its topic). Existing models predicting retweets focus almost exclusively on the tweet's structural elements (e.g., Petrovic, Osborne & Lavrenko, 2011; Suh et al., 2010), because this information is easily extractable from Twitter. Those studies looking at how content predicts retweeting are based on the automated derivation of content-related variables rather than on researcher interpretation. For instance, Uysal & Croft (2011) look at tweet novelty by calculating the lexicographical similarity between a tweet and its nearest neighbors in users' timelines.

Regarding the structural elements of a tweet, the evidence on the role that hashtags and URLs play in determining retweets is mixed. For example, Suh et al. (2010) and boyd et al. (2010) both find that including hashtags or URLs in tweets boosts their chances of getting retweeted. Conversely, Liu et al. (2012) find that including a URL in a tweet affects its

[2] The extent to which an MP's seat is considered to be marginal, calculated according to Finer, Berrington, & Bartholomew's (1961) model of marginality as cited in Jackson and Lilleker (2011) and used in their research. Categories are safe (a majority of 11% of votes over the next nearest candidate), near-marginal (between 5.1% and 10.9%), and marginal (less than 5%).

likelihood of being retweeted negatively, although including multimedia such as pictures or videos has a positive impact. However, their research was done in the specific context of retweeting in public emergencies. In a commercial marketing context, Malhotra, Malhotra, & See (2012) suggest that neither URLs nor hashtags increase retweets. The jury remains out on the effect of hashtags and URLs on retweeting.

Research on whether people are more likely to disseminate positive or negative information provides mixed conclusions (King, Racherla, & Bush, 2014). Some indicate positive news content obtains greater dissemination (Berger & Milkman, 2012) whilst others find that negative news content is more transmissible (Hansen et al., 2011). It is therefore reasonable to assume that a tweet's sentiment– i.e., whether it is positive or negative - is likely to influence the likelihood of its retweeting. Kupavskii et al. (2012) include sentiment considerations in their retweet prediction model, extrapolating sentiment automatically based on the appearance of positive or negative terms and smileys in each tweet. However, their approach neglects many of the subtleties of sentiment (e.g. sarcasm) that a human coder would spot. Lemahieu et al. (2015) examine which tweets were most likely to be retweeted from 1,000 tweets by 500 top Twitter users in eight different categories. They find that tweet sentiment was the least predictive of the factors they considered. However, their sentiment score was also automatically-computed and measured the overall strength of sentiment in a tweet – its emotionality - rather than the valence of the sentiment expressed. In a study looking at retweeting in the run-up to the 2011 Berlin state parliamentary election, Dang-Xuan et al. (2013) consider both the level of emotionality in the tweet and its content. The higher the level of emotionality, the more likely retweeting occurred. They also found politicians were much less likely to include emotionality in their tweets, focusing instead on information transmission as opposed to opinion. In contrast, private individuals tended to use tweets to diffuse subjective and emotional information. Messages containing some kind of

appraisal of politicians/ political parties were more likely to be retweeted, suggesting that the topic of the tweet plays a role in determining retweets. There is a considerable corpus of literature examining the content and purpose of politicians' tweets (e.g., Tumasjan et al., 2011; Golbeck, Grimes, & Rogers, 2010; Hemphill, Otterbacher, & Shapiro, 2013), but such research focuses more on topic categorization rather than identifying what different types of message content drive retweeting.

Figure 1 and Table 1 summarize the factors discussed above which act as potential antecedents to retweeting behavior, categorized under source (sender) characteristics and message (structural elements, sentiment and content) characteristics.

***** Insert Figure 1 and Table 1 about here *****

METHODOLOGY

This study develops a methodological process for the prediction of retweeting behavior (see Figure 2). It builds predictive models for the likelihood of message retweeting based on predictors including three forms of message characteristics and one form of source characteristic: (i) tweet structural elements, (ii) tweet content and (iii) sentiment of tweets; and (iv) source characteristics.

The unit of analysis in this study is taken to be the individual tweet. Twitter automatically appends structural data to each tweet. Some sender data are available from Twitter (e.g. number of followers, number of tweets sent) and other data can be manually appended (e.g. gender, party affiliation, age). Tweet message content and sentiment is obtained through manual coding.

Sample

The sample consists of 42,444 original tweets sent by 366 British MPs during the 2015 UK General Election campaign running from the dissolution of Parliament on 20 March to polling day on 7 May. The tweets were collected using Brandwatch, a commercially-available social media listening tool enabling access to all sent tweets and their associated metadata. A list of all MPs on Twitter is maintained by the Twitter account @tweetminister and this was used as the basis of the data collection, supplemented by manual checking to ensure list accuracy. All tweets analyzed were original posts generated by the MPs themselves, forming part of the way they wished to present themselves during an election campaign. Any retweets or replies to other people by MPs were excluded from the analysis.

Method

Several stages of analysis were undertaken on different sections of the data. The first stage consisted of a descriptive analysis to determine what politicians were tweeting, how frequently they were tweeting and the nature of the campaign tweets. This analysis was undertaken on the entire tweet population. The second stage included predictive models to ascertain the extent to which the structural elements of the tweet and the sender characteristics influenced the likelihood of a tweet being retweeted. Of the 42,444 original tweets, 6,510 were not retweeted (15.3%). These 6,510 were matched with a random sample of 6,510 tweets which were retweeted to produce an evenly matched sample of 13,020 tweets on which to conduct the predictive analysis for source characteristics and tweet structural data. Such a procedure was necessary otherwise a predictive model could have been correct in its predictions 84.7% of the time simply by predicting that every tweet would be retweeted. The last stage consisted of manual sentiment and content analyses of the tweets, used in predictive models of retweeting. This stage involved a reduced sample of 1,212 tweets (again evenly split between messages that were/ were not retweeted).

***** Insert Figure 2 about here *****

Each of these tweets was manually coded by one of the researchers for sentiment and content. Firstly, each tweet was read and coded as positive, negative or neutral. The file of tweets to be manually coded included the tweet content, the sender's name (knowledge of the tweet source was often required to ascertain whether a positive or negative message was intended) and a unique reference number to enable the sample to be reintegrated into the main dataset. No information about whether or not each tweet was retweeted was available as the manual coding was undertaken, to avoid coding bias.

Positive tweets included some positive sentiment, e.g. a statement of achievement, thanking people, offering support to someone else, and reporting good news. Negative tweets criticized the opposition, for example, or planted fear/concern in people's minds regarding the opposition winning and use of sarcasm to imply criticism. Neutral tweets did not include any emotional element, e.g. those simply passing on information or describing something without additional commentary. Each tweet could only have one code, i.e. positive or negative, not both. Where a tweet contained both positive and negative sentiment, the researcher decided which element dominated and coded accordingly.

The researcher then coded the tweets for content. The tweet content codebook developed was informed by frameworks using existing approaches for categorizing political tweet types (e.g. Tumasjan et al., 2011; Golbeck, Grimes, & Rogers, 2010; Hemphill, Otterbacher, & Shapiro, 2013). However, existing coding schema either relate to an American context (Golbeck, Grimes, & Rogers, 2010; Hemphill, Otterbacher, & Shapiro, 2013) or have been developed for categorizing non-campaigning tweets (Jackson & Lilleker, 2011). Additionally, tweeting behavior has changed over time and schemas developed several years ago needed updating to account for how MPs are using Twitter. During the coding, it

became apparent that certain tweets needed new codes, e.g. charity tweets in which MPs express support for a certain charity.

In line with Hemphill, Otterbacher, & Shapiro (2013), each tweet was coded as many times as was relevant. Although tweets are short (140 characters), it is possible to pack considerable information into them and MPs frequently send tweets for multiple communication purposes. A sample of 15% of the coded tweets was then coded by a second independent coder. Intercoder agreement on the sentiment of tweets was 87.8%. Agreement on the content codes was in excess of 90% for all content-related codes. Generally a minimum intercoder reliability of 80% is seen as desirable (Riffe, Lacy and Fico, 2014). A summary of the coding scheme, along with how the codes relate to the extant literature and the number of times each code appeared in the data sample is illustrated in Table 2.

***** Insert Table 2 about here *****

Measures

Dependent variable: Retweeting

The dependent variable in this study is whether a message is retweeted or not. The fact that someone retweets a message is taken as a measure of sharing and engaging with that message, as per other studies measuring engagement and influence on Twitter (e.g. Lotan et al., 2011; Araujo, Neijens, & Vliengenthart, 2015).

Independent variables

The analysis employed independent variables including source (sender) characteristics – in this context the member of parliament (MP) sending the Twitter message - and tweet message structural data, and variables newly coded through sentiment and content analysis of the tweeted messages. How each of these variables was derived is illustrated in Table 1.

Analysis: predictive model

The aim of this research is to build a series of classification models to predict retweeting. The models are built using CHAID, i.e., chi-square automatic interaction detector (Kass 1980). CHAID is a rule induction modelling technique. It proposes a set of rules used to classify groups within the data as they relate to the target variable. It is often referred to as tree analysis because a target variable node (tree trunk) is split into predictor nodes (branches). The usage of CHAID was driven by both the nature of the data - CHAID can accommodate non-parametric data and does not assume data to be normally distributed - and by the need to derive actionable insight. The resulting model provides politicians and marketers with workable rules to maximize the probability of a message being retweeted.

CHAID works by building a decision tree using significance testing (in this case using chi-square tests as the outcome being predicted is categorical) to identify which factors are relevant in predicting a particular outcome. Before modelling commenced, the data were partitioned into a training set and a validation set. Each model was built using the training data and then tested on the validation data. The CHAID algorithm evaluates each predictor in turn and selects whichever splits the data best into two groups (retweeted or not) based on the value to be predicted. Its aim is to identify the two purest subgroups possible in relation to the outcome of interest. Thus one group will have as many retweeted tweets as possible and the other will have as many non-retweeted tweets. The algorithm considers each of the two groups and further splits them according to the same principles. This iterates until no further subgroups can be usefully formed (or the maximum number of splits the analyst specified is reached). IBM SPSS Modeler was used to run the CHAID models.

CHAID has been used previously in the political marketing literature, for example, to identify voter segments (Baines et al., 2003), and in market segmentation generally (Haughton & Oulabi, 1997; McCarty & Hastak, 2007; Hsu & Kang, 2007), as well as to

categorize what factors lead to particular outcomes (e.g. in health research on trauma outcomes: Hill, Delaney & Roncal, 1997). However, it does not appear to have been previously used to predict retweeting or, in particular, to identify those factors that are most relevant to stimulating retweeting. Indeed, previous retweeting research (e.g. Mahmud, Chen, & Nichols, 2014, Lemahieu et al., 2015, Hou, Huang, & Zhang, 2015) has tended to treat retweet prediction as a computational challenge, focusing almost exclusively on generating as accurate a prediction score as possible without emphasis on identifying the factors which drive retweeting. However, for marketers, knowledge of factors driving retweeting is as important, if not more so, than generating an accurate prediction score because knowledge of the factors that drive the score can be used to positively influence marketing practice. Therefore, it is critical that whatever predictive approach is used makes it possible for these factors to be easily identified.

In this regard, CHAID offers substantial advantages for marketers over other predictive techniques such as neural networks and logistic regression because it is a rule induction approach, meaning that the way in which it derives its prediction is expressed in the form of a series of rules that a marketer can follow to improve the chances of their own tweets being retweeted. This is not the case with neural networks, which use a network of hidden connections to draw their conclusions and do not tell the analyst how the score is derived (Struhl, 2015) or logistic regression which produces an equation which can be used to predict outcomes but which cannot easily be translated into a set of business rules to inform practice. Another advantage of CHAID is that decision trees are less affected by extreme values than either neural networks or logistic regression. This makes CHAID particularly well-suited to the analysis of Twitter data as many variables related to Twitter tend to have highly skewed long-tail Poisson distributions with many extreme values (e.g. number of followers, volume of tweets sent, number of retweets per tweet).

FINDINGS

Four CHAID models were built[3]. Model 1 used only variables relating to the tweets' structural elements. Model 2 used only variables relating to tweets' senders. Model 3 used content and sentiment variables only and Model 4 combined all possible variables. Next, the results including the outcome of the manual sentiment and content analyses are described.

CHAID Model 1: Structural elements of the tweet only

The first CHAID model used only variables relating to the tweets' structural characteristics. These variables included the time of sending, whether the tweet incorporated a hashtag (y/n), the number of hashtags included, the number of @mentions in the message, whether it included a URL, and whether it included a link to a picture or video (i.e., media link). The resultant CHAID model determined that the tweet sending time was not relevant and excluded it from the model. All other variables were included and the relative importance of each in the overall model is described in Table 3.

Overall, Model 1 was correct in its predictions only 59.29% of the time. However, within the model, there were some individual nodes in which the chances of a retweet were considerably higher. For example, tweets with at least one hashtag, at least one @mention and at least one media link were retweeted 74% of the time. Conversely, tweets precluding hashtags or links but which mentioned one or more other people were only retweeted 33% of the time.

***** Insert Table 3 about here *****

CHAID Model 2: Source (sender) characteristics only

[3] Given time and space constraints, the CHAID computer output for Models 1, 2, 3 and 4 is not outlined here but these are available from the corresponding author. Please contact Paul Baines - contact details are made available at the end of the article - if you require these.

Source characteristics entered into the analysis included the number of people followed on Twitter (followees), the number of Twitter followers, the ratio of followers to followees, the total number of tweets ever sent on Twitter, the total number of campaign tweets sent, the number of tweets sent per day of the campaign, the gender, the age, the senders' party affiliation, their percentage majority before the election, the marginality of their seat and the year they entered parliament.

The first iteration of Model 2 correctly predicted whether or not a tweet would be retweeted 72.4% of the time. By far the most important variable in determining whether a tweet is retweeted or not is the number of followers a sender has. The more followers someone has, the more people see a tweet and hence the more likely it is that someone will retweet it. However, the sender of the tweet does not have direct control over the number of people following them, nor over variables such as age, gender, marginality of seat, party affiliation, year entered parliament or their percentage majority before the election. To improve managerial relevance and provide practical recommendations to MPs on how to improve their retweet rates, the model was run again with all uncontrollable variables removed. The second iteration of this model used only the number of people the MP follows, the ratio of followers to followees, the total number of tweets they sent, the total number of campaign tweets they sent and the mean number of campaign tweets per day. This revised model was accurate in its predictions 69.36% of the time.

Compared to the previous tweet structural elements analysis, the characteristics of the tweet sender are better predictors of retweeting than the tweet structural elements. Those message source variables with the most predictive power are included in Table 3.

Sentiment analysis

Out of 1,212 tweets coded, 16.74% were deemed to be negative, 46.78% neutral and 36.36% positive. As Table 4 shows, the sentiment of the tweets has a very clear impact on how likely they are to be retweeted. Almost 80% of the negative tweets were retweeted, compared to just over 50% of the positive ones (79.8% vs. 50.9%, $\chi^2 = 96.731$, $p<.01$). For neutral tweets the rate of retweeting fell to just under 40%. These results suggest that negative tweets are much more likely to be retweeted than are either positive or neutral tweets.

***** Insert Table 4 about here *****

Content analysis

Chi square tests were run for each of the coded categories to determine whether there was a significant relationship between the content/topic of the tweet and whether or not it was retweeted. The results of these tests are outlined in Table 5. Six of the content variables were positively associated with retweeting and two were negatively associated with retweeting (significant variables are highlighted using asterisks), suggesting that the content of the tweet can influence the chances of it being retweeted.

***** Insert Table 5 about here *****

CHAID Model 3: Message content and sentiment codes only

All the content and sentiment variables were used to build another CHAID model. This new model, Model 3, indicated sentiment to be the most powerful predictor of retweeting, with 77.4% of negative tweets being retweeted. For positive tweets to be retweeted, the best thing an MP can do is call for people to vote (100% retweeted). Message content referring to the campaign trail that combines an expression of support for someone else is also retweeted 100% of the time.

Model 3 was correct in its predictions 63.31% of the time (Table 3). For the same model run again with sentiment excluded, the most important content-related predictor of retweeting becomes whether or not a tweet is an attack tweet. 69.34% of attack tweets are retweeted. Within this category, an MP can boost the chances of a retweet further by writing a tweet containing a fear appeal. Almost 90% of attack fear appeal tweets were retweeted. These two variables are not used in the model if sentiment is included because almost all attacking and fear appeal tweets have also been coded as negative in sentiment, and so the model does not need to include all three variables to be predictive.

CHAID Model 4: Combined source, message structure, message content and message sentiment variables

A final model incorporated all the variables so far discussed – the Twitter structural data, the senders' data and the content and sentiment-related data (including those variables excluded from Model two because the tweet senders could not control them) –to evaluate how these factors interact to predict retweeting. This model had a predictive accuracy of 71.67%. The number of Twitter followers of the individual emerged as the single most important predictor of retweeting, followed by the sentiment of the tweet. The relative importance of the most predictive variables is shown in Table 3.

Those tweets from MPs with fewer than 4,079 followers were only retweeted in 23.89% of cases: they could do little to improve their retweeting chances other than by trying to obtain more followers. Those MPs with between 4,079 and 11,719 followers had their messages retweeted in 42.75% of cases. Retweets rose to 64.11% for MPs with between 11,720 and 32,135 followers and up to 80.72% for MPs with more than 32,135 followers. At the top end of the scale, MPs with more than 321,135 followers will find almost all their tweets retweeted just by virtue of the number of followers they have.

DISCUSSION

This study makes both a methodological and theoretical contribution by deriving a model to predict the likelihood of Twitter messages being retweeted in political marketing election campaigns. The variables used to predict message retweet likelihood included message characteristics (a) tweet structural elements, b) content of tweets and c) sentiment of tweets, and source characteristics including d) the characteristics of the tweet sender.

In addition to manual sentiment and content analysis coding, the analytical approach consisted of running different predictive CHAID decision tree models that either separately used the different categories of predictors or considered these predictors in an aggregated fashion. To the best knowledge of the authors, this approach has not been used to predict retweeting previously. Figure 2 presents the full methodological approach adopted in this study. This methodological approach is replicable in other (i.e. commercial) settings. Model 1, using the structural elements of the tweet, predicted retweeting correctly on 59.29% of occasions. While this level of predictive power is relatively low, the analysis identified that engineering certain elements of a Twitter message can yield a higher likelihood of retweeting, including: incorporating a hashtag, a mention and a media link improves retweeting rates to 74.13%. Comparatively, Model 2 - the sender-characteristics predictive model - predicted retweeting correctly on 72.4% of occasions. Of these characteristics, the sender's number of Twitter followers was the strongest predictor of retweeting.

The manual coding of tweets identified variables that predicted the likelihood of retweeting. Based on the coding of sentiment, the finding that negative tweets are more impactful than positive tweets parallels findings obtained elsewhere (see Hennig-Thurau, Wiertz, & Feldhaus, 2014). The manual classification of tweets revealed that negative messages were retweeted on 79.8% of occasions.

The manual coding of messages' content also revealed categories that have not been identified in previous research, including political Twitter messages classified as 'attack' or 'fear appeals'. The types of tweets that distinguish themselves as being retweeted included attack tweets (73.6%), support for others' tweets (77.3%) and fear appeal tweets (89.7%). Conversely, tweets of a personal nature had a low degree of retweeting (30.9%). The CHAID model, using content-based categories, revealed that retweeting is best predicted by whether or not a tweet is of the 'attack' type. An attack tweet, also including a fear appeal, had an even higher likelihood of being retweeted, at 90%.

Model 3 based on all message characteristics revealed that the tweet's sentiment was the strongest predictor of retweeting. However, even though followers generally retweet positive messages less than negative messages, positive tweets do garner very high retweeting rates depending on additional message characteristics: for instance, the positive tweets that mention support for others (a retweeting rate of 76%).

Finally, Model 4, comprising all the variables described above, had the highest predictive power of 76.29%, with the strongest predictor being the number of Twitter followers followed by the tweet's sentiment. This model extends current understanding of the antecedents of tweeting behavior, thereby making a theoretical contribution to the eWOM literature. Whereas, the number of Twitter followers is not easily adjusted/upgraded (at least not in the short term), the tweet valance can be adjusted easily. MPs in the lowest group can do little to improve their retweeting rate beyond getting more followers as the only other variables that have an influence in this group – percentage majority before the election and year entered parliament – are things they cannot control. However, once an MP gains more than 4,079 followers, they can increase their chances of getting retweeted up to 70.9% by going negative significantly or by expressing support for others (85%).

Theoretical and Practitioner Contributions

Contribution to Theory

There are few published studies of retweeting behavior, especially in marketing journals. Most current evidence relates to articles submitted to computing science conferences and workshops. This work makes a contextual contribution to the literature on retweeting behavior in a (political) marketing context but also a methodological contribution by highlighting a proven method to ascertain what the antecedents of retweeting behavior are. The findings correspond with those of Morchid et al. (2014) who considered the volume (rather than likelihood) of retweeting. They used principal components analysis to identify massively retweeted tweets (with a classification accuracy of 65.7%), and found that that this category of tweet tends to incorporate the following five features: (i) number of hashtag topics in a tweet; ii) number of followers; (iii) mentions (number of cited usernames); iv) status (number of previous tweets written) and v) number of contained URLs.

Previous studies of Twitter use in political marketing have identified that Twitter users are more politically engaged and less trusting of mainstream media and that having a large number of followers is important in retweeting but that these followers need encouragement to retweet (Bode & Dalrymple, 2015). However, to the best knowledge of the authors, previous work is yet to identify the characteristics of retweeted messages using a combination of *all* source and message effects including content and sentiment, as undertaken here.

The findings indicate that negative tweets are more likely to be retweeted, in the political context, than positive tweets. This finding is contrary to some extant studies in the WOM literature (see East, Hammond, & Wright, 2007; Wu, 2013) which have found positive word of mouth is more transmissible. This finding therefore deserves further investigation to

ascertain its replicability in this and other settings. It may be that this is a category effect and that political retweeting differs from that concerning say FMCG products.

Contribution to Practice

The findings of this study are useful not only for political marketers but also for marketers more generally. Of particular importance to marketers is the notion of how to develop a process to determine what characteristics are important in increasing the transmissibility of Twitter marketing messages generally, and specifically promoted tweets (e.g. adverts), messages aimed at managing a crisis (e.g. a product recall or ethical crisis), mass customer service messaging and thought leadership messaging (e.g., IBM's Smarter Planet initiative).

For political marketers, the findings indicate that developing a large network of influential followers is fundamental to generating a high retweet rate and that some messages are more likely to be retweeted if they contain negative, attack or fear appeal content. Campaigners are, however, advised to extend caution here, as the wrong kind of 'tone' can elicit public outrage with potentially significant outcomes. In the political marketing setting, there is a balance to be struck between using an emotive appeal to cut through the Twitter clutter and using an appeal that is appropriate for the circumstances.

Limitations and Further Research

Further research should be conducted to identify whether or not the patterns identified in this study are replicated in other settings. For example, whilst this study considered the retweeting of tweets sent out by British politicians in a general election setting (a person-person interrelationship), retweeting behavior may differ in other political contexts (e.g. the US) or when tweets are sent out by organizations, e.g. a referendum, single issue or party

campaign group (i.e. organization-person interrelationship), or individual companies, or for promoted tweets since the relationship between followers and a group operate differently. The political context also allows the analysis of the operation of multiple 'marketers' at any one time, but there are likely to boundaries to the application of the current findings to commercial marketing contexts where individual receivers are exposed to messages from commercial organizations. The answers to this question are outside the scope of this study. Further research is necessary to understand the antecedents of retweeting behavior in this setting and in what types of relationships.

Our study did not consider who is doing the retweeting and to whom they are retweeting. This is important in a political marketing setting aimed at understanding the influence of retweeting because one can differentiate between party activists sending tweets to other activists versus switch voters sending tweets to other potential switch voters.

Concluding Remarks

This study highlights how a series of CHAID analysis models using readily available Twitter data and manually coded sentiment and content data were developed to identify why some tweets are more likely to be retweeted when compared to others in the context of a UK political campaign. Four models were developed with different rates of predictive accuracy including: 1) using message structure variables only (59.29% predictive accuracy); 2) tweet source characteristics only (72.4% accuracy); 3) content and sentiment analysis variables only (64.43% accuracy) and 4) combined source, message structure, message content and message sentiment (76.29% accuracy). The final model (Model 4) indicates that the strongest predictor of retweeting is the source's number of Twitter followers followed by the tweet's sentiment. Whereas the number of Twitter followers is not easily affected, tweet valance can be adjusted more easily. However, once an MP gains more than 4,079 followers,

they can increase their chances of getting retweeted up to 70.9% by going negative or by expressing support for others. Finally, this work represents a call to arms to other (political) marketing researchers to use the current methodological approach, or derivations of it, to further understanding of retweeting behavior in other settings.

REFERENCES

Agranoff, C., & Tabin, H. (2011). Socially elected: How to win elections using social media. West Jordan, UT: Pendant Publishing.

Araujo, T., Neijens, P., & Vliegenthart, R. (2015). What motivates consumers to re-tweet brand content? The impact of information, emotion, and traceability on pass-along behavior. Journal of Advertising Research, DOI: 10.2501/JAR-2015-009.

Baines, P. R., Worcester, R. M., Jarrett, D., & Mortimore, R. (2003). Market segmentation and product differentiation in political campaigns: A technical feature perspective. Journal of Marketing Management, 19(1-2), 225-249.

Bakshy, E., Hofman, J., Mason, W., & Watts, D. (2011). Everyone's an influencer: quantifying influence on Twitter. Proceedings of the fourth ACM international conference on web search and data mining. SE-WSDM '11 (pp. 65–74). Hong Kong. DOI: 10.1145/1935826.1935845.

Berlo, D. K., Lemert, J. B., & Mertz, R. J. (1969). Dimensions for evaluating the acceptability of message sources. Public Opinion Quarterly, 33(4), 563-576.

Bode, L., & Dalrymple, K. E. (2014). Politics in 140 characters or less: Campaign communication, network interaction, and political participation on Twitter. Journal of Political Marketing. DOI: 10.1080/15377857.2014.959686.

Berger, J. & Milkman, K. L. (2012). What makes online content viral? Journal of Marketing Research, XLIX(April), 192–205.

boyd, d., Golder, S., & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. Proceedings of the Annual Hawaii International Conference on System Sciences, 1–10. DOI: 10.1109/HICSS.2010.412.

Campanile, C. (2013). Weiner caught sending dirty messages and photos a year after his sexting scandal. New York Post. 24 July. Retrieved 7 February, 2016, from:

http://nypost.com/2013/07/24/weiner-caught-sending-dirty-messages-and-photos-a-year-after-his-sexting-scandal/.

Cha, M., Haddadi, H., Benevenuto, F., & Gummadi, K. P. (2010). Measuring user influence in Twitter: The million follower fallacy. Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media. Retrieved 29 May, 2016, from: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1538/1826,2011.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. Journal of Marketing Research, 43(3), 345-354.

Chiu, H., Hsieh, Y., Kao, Y., & Lee, M. (2007). The determinants of email receivers' disseminating behaviors on the internet. Journal of Advertising Research, 47(4), 524-534.

Chiu, H., Pant, A., Hsieh, Y., Lee, M., Hsioa, Y., & Roan, J. (2014). Snowball to avalanche. European Journal of Marketing, 48(7/8), 1255-1273.

Cialdini, R. B. (1993). Influence: The psychology of persuasion. New York: William Morrow & Co.

Cook, L. (2013). How is social media changing politics? In Friends of Europe. Debating Europe, 22 January. Retrieved 29 May, 2016, from: http://www.debatingeurope.eu/2013/01/22/how-is-social-media-changing-politics/#.V0rYo_krLIU.

Dang-Xuan, L., Stieglitz, S., Wladarsch, J., & Neuberger, C. (2013). An investigation of influentials and the role of sentiment in political communication on Twitter during election periods. Information, Communication & Society, 16(5), 795–825.

Davies, R. (2014). Social media in election campaigning. European Parliamentary Research Service Briefing, 21 March. Retrieved 29 May, 2016, from: http://www.europarl.europa.eu/RegData/bibliotheque/briefing/2014/140709/LDM_BRI(

2014)140709_REV1_EN.pdf.

East, R., Hammond, K., & Wright, M. (2007). The relative incidence of positive and negative word of mouth: A multi-category study. International Journal of Research in Marketing, 24(2), 175-184.

Eagly, H., & Chaiken, S. (1993). The psychology of attitudes, Fort Worth, TX: Harcourt Brace Jovanovich.

Enli, G. S., & Skogerbø, E. (2013). Personalized campaigns in party-centred politics. Information, Communication & Society, 16(3), 757–774.

Finer, S.; Berrington, H., & Bartholomew, D. (1961). Backbench opinion in the House of Commons, 1955–59. New York: Pergamon.

Gilly, M. C., Graham, J. L., Wolfinbarger, M. F., & Yale, L. J. (1998). A dyadic study of interpersonal information search. Journal of the Academy of Marketing Science, 26(2), 83-100.

Golbeck, J., Grimes, J. M., & Rogers, A. (2010). Twitter use by the US Congress. Journal of the American Society for Information Science and Technology, 61(8), 1612–1621.

Graham, T., Broersma, M., & Hazehoff, K. (2013). Closing the gap? Twitter as an instrument for connected representation. In R. Scullion, R. Gerodimos, D. Jackson, & D. Lilleker (Eds.). The media, political participation and empowerment (pp.71-88). Oxford: Routledge.

Graham, T., Broersma, M., Hazelhoff, K., & van't Haar, G. (2013). Between broadcasting political messages and interacting with voters: The use of Twitter during the 2010 UK general election campaign. Information, Communication & Society, 16(5), 692–716.

Hansen, L. K., Arvidsson, A., Nielsen, F. A., Colleoni, E., & Etter, M. (2011). Good friends, bad news: Affect and virality in Twitter. Proceedings of the International Workshop on Social Computing, Network, and Services (pp. 1–14). Crete, Greece.

Haughton, D., & Oulabi, S. (1997). Direct marketing modeling with CART and CHAID. Journal of Interactive Marketing, 11(4), 42-52.

Helm, T. (2013). Obama's tech wizard hones his media weapons to help Miliband get to No 10. The Guardian. Retrieved 30 November, 2013, from: http://www.theguardian.com/politics/2013/nov/30/obama-miliband-no-10-mitt-romney-labour-campaign.

Hemphill, L., Otterbacher, J., & Shapiro, M. (2013). What's Congress doing on Twitter? Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13), 877. DOI: 10.1145/2441776.2441876.

Hennig-Thurau, T., Wiertz, C., & Feldhaus, F. (2014). Does Twitter matter? The impact of microblogging word of mouth on consumers' adoption of new movies. Journal of the Academy of Marketing Science, 43 (3), 375–394.

Hermida, A., Lewis, S. C., & Zamith, R. (2014). Sourcing the Arab Spring: a case study of Andy Carvin's sources on Twitter during the Tunisian and Egyptian revolutions. Journal of Computer-Mediated Communication, 19(3), 479-499.

Hill, D. A., Delaney, L. M., & Roncal, S. (1997). A chi-square automatic interaction detection (CHAID) analysis of factors determining trauma outcomes. Journal of Trauma and Acute Care Surgery, 42(1), 62-66.

Hovland, C.I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. Public Opinion Quarterly, 15(4), 635-50.

Hsu, C. H., & Kang, S. K. (2007). CHAID-based segmentation: International visitors' trip characteristics and perceptions. Journal of Travel Research, 46(2), 207-216.

internetlivestats (2016). Twitter usage statistics. Retrieved 14 August, 2016, from: http://www.internetlivestats.com/twitter-statistics/#trend.

Jackson, N., & Lilleker, D. (2011). Microblogging, constituency service and impression

management: UK MPs and the use of Twitter. The Journal of Legislative Studies, 17(1), 86–105.

Jansen, B. J., & Zhang, M. (2009). Twitter power: Tweets as electronic word of mouth. Journal of the American Society for Information Science and Technology, 60(11), 2169–2188.

Jin, S.-A. A., & Phua, J. (2014). Following celebrities' tweets about brands: The impact of Twitter-based electronic word-of-mouth on consumers' source credibility perception, buying intention, and social identification with celebrities. Journal of Advertising, 43(2), 181–195.

Kahle, L. R., & Homer, P. M. (1985). Physical attractiveness of the celebrity endorser: A social adaptation perspective. Journal of Consumer Research, 11(4), 954-961.

Kaid, L. L. (2002). Political advertising and information seeking: Comparing exposure via traditional and Internet channels. Journal of Advertising, 31(1), 27-35.

Kamins, M. A. (1989). Celebrity and noncelebrity advertising in a two-sided context. Journal of Advertising Research, 29(3), 34-42.

Karmarkar, U. R., & Tormala , Z. L. (2010). Believe me, I have no idea what I'm talking about: The effects of source certainty on consumer involvement and persuasion. Journal of Consumer Research, 36(6), 1033-104.

Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. Journal of the Royal Statistical Society Series C (Applied Statistics), 29(2), 119-127.

Kim, E., Sung, Y., & Kang, H. (2014). Brand followers' retweeting behavior on Twitter: How brand relationships influence brand electronic word-of-mouth. Computers in Human Behavior, 37(February), 18–25.

King, R. A., Racherla, P., & Bush, V. D. (2014). What we know and don't know about online

word-of-mouth: A review and synthesis of the literature. Journal of Interactive Marketing, 28(3), 167–183.

Kirmani, A., & Shiv, B. (1998). Effects of source congruity on brand attitudes and beliefs: The moderating role of issue-relevant elaboration. Journal of Consumer Psychology, 7(1), 25-47

Kupavskii, A., Ostroumova, L., Umnov, A., Usachev, S., Serdyukov, P., Gusev, G. & Kustarev, A. (2012). Prediction of retweet cascade size over time. Proceedings of the 21st ACM International Conference on Information and Knowledge Management, (February 2016), 2335–2338. DOI: 10.1145/2396761.2398634.

Larsson, A. O. & Hallvard, M. (2011). Who tweets? Tracking microblogging use in the 2010 Swedish election campaign. Proceedings of European Conference on Information Ssystems (ECIS) Paper 251. Retrieved 29 May, 2016, from: http://aisel.aisnet.org/ecis2011/251/.

Larsson, A. O., & Kalsnes, B. (2014). 'Of course we are on Facebook': Use and non-use of social media among Swedish and Norwegian politicians. European Journal of Communication, 29(6), 653-667.

Lawless, J. L. (2012). Twitter and Facebook: New ways for members of Congress to send the same old messages? In R. L. Fox & J. M. Ramos (Eds.), Citizens, elections and governing in the new media era (pp. 206–232). Cambridge: Cambridge University Press.

Lee, B. (2014). Window dressing 2.0: Constituency-level web campaigns in the 2010 UK general election. Politics, 34(1), 45–57.

Lee, M., & Youn, S. (2009). Electronic word of mouth (eWOM). International Journal of Advertising, 28(3), 473-499.

Lemahieu, R., Van Canneyt, S., De Boom, C., & Dhoedt, B. (2015). Optimizing the popularity of Twitter messages through user categories. 2015 IEEE International

Conference on Data Mining Workshop (ICDMW) (pp. 1396–1401). DOI: 10.1109/ICDMW.2015.39.

Liu, Z., Liu, L., & Li, H. (2012). Determinants of information retweeting in microblogging. *Internet Research*, *22*(4), 443-466.

Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I., & boyd, d. (2011). The revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. International Journal of Communications, 5, 1375–1406.

Malhotra, A., Malhotra, C. K., & See, A. (2012). How to get your messages retweeted. MIT Sloan Management Review, 53(Winter), 61–66.

McCarty, J. A., & Hastak, M. (2007). Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. Journal of Business Research, 60(6), 656-662.

MacDonald, J., & Elahee, M. N. (2003). Examining the relationship between website visits and purchase intentions for internet shoppers: implications for international marketers. Journal of International Business and Entrepreneurship Development, 1(2), 69-75.

Mahmud, J., Chen, J., & Nichols, J. (2014). Why are you more engaged? Predicting social engagement from word use, Computing Research Repository. Retrieved 6 November, 2016, from http://arxiv.org/abs/1402.6690.

Mitchell, A., Holcomb, J., & Weisel, R. (2016) Election 2016: Campaigns as a Direct Source of News, Pew Research Center, 18 July. Retrieved 2 September, 2016, from: http://www.journalism.org/2016/07/18/election-2016-campaigns-as-a-direct-source-of-news/.

Morchid, M., Dufour, R., Bousquet, P. M., Linarès, G., & Torres-Moreno, J. M. (2014). Feature selection using Principal Component Analysis for massive retweet detection. Pattern Recognition Letters, 49, 33-39.

Newman, N. (2010) #UKelection 2010, mainstream media and the role of the Internet: How social and digital media affected the business of politics and journalism. Retrieved 6 November, 2016, from http://reutersinstitute.politics.ox.ac.uk/sites/default/files/Social Media and the Election.pdf.

Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. Journal of Advertising, 19(3), 39-52.

Park, C., & Lee, T. M. (2009). Antecedents of online reviews' usage and purchase influence: An empirical comparison of U.S. and Korean consumers. Journal of Interactive Marketing, 23(4), 332-340.

Park, D., & Kim, S. (2008). The effects of consumer knowledge on message processing of electronic word-of-mouth via online consumer reviews. Electronic Commerce Research & Applications, 7(4), 399-410.

Petrovic, S., Osborne, M., & Lavrenko, V. (2011). RT to win! Predicting message propagation in Twitter. Proceedings of the Fifth International Conference on Weblogs and Social Media (ICWSM '11) (pp. 586–589). Retrieved 29 May, 2016, from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2754.

Petty, R. E., & Cacioppo, J. T. (1984). Source factors and the Elaboration Likelihood Model of persuasion. Advances in Consumer Research, 11(1), 668-672.

Petty, R. E., Cacioppo, J. T., & Schumann, D. (1983). Central and peripheral routes to advertising effectiveness: The moderating role of involvement. Journal of Consumer Research, 10(2), 135-146.

Riffe, D., Lacy, S. & Fico, F. (2014). Analysing media messages: Using quantitative content analysis in research (3e). Abingdon: Routledge.

Rogers, E. (2003). Diffusion of innovations (5e). New York: The Free Press.

Ross, K., & Burger, T. (2014). Face to Face(book): Social media, political campaigning and the unbearable lightness of being there. Political Science, 66(1), 46–62.

Rudat, A., Buder, J., & Hesse, F. W. (2014). Audience design in Twitter: Retweeting behavior between informational value and followers' interests. Computers in Human Behavior, 35, 132–139.

Sæbø, Ø. (2011, August). Understanding Twitter™ use among parliament representatives: A genre analysis. In International Conference on Electronic Participation (pp. 1-12). Berlin: Springer.

Shannon, C.E., & Weaver, W. (1949). The mathematical theory of communication. Urbana, IL: University of Illinois Press.

Struhl, S. (2015). Practical text analytics. London: Kogan Page.

Suh, B., Hong, L., Pirolli, P., & Chi, E. H. (2010). Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. Social Computing (SocialCom), 2010 IEEE Second International Conference (pp. 177–184). Minneapolis.

Tang, X., Miao, Q., Quan, Y., Tang, J., & Deng, K. (2015). Predicting individual retweet behavior by user similarity: A multi-task learning approach. Knowledge-Based Systems, 89, 681–688.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2011). Election forecasts with Twitter: How 140 characters reflect the political landscape. Social Science Computer Review, 29(4), 402–418.

Tumulty, K. (2007). Obama's viral marketing campaign. Time. 5 July. Retrieved 7 February, 2016, from http://content.time.com/time/magazine/article/0,9171,1640402,00.html.

Twitter (2016). Twitter usage: company facts. Twitter, 30 June. Retrieved 14 August, 2016, from https://about.twitter.com/company.

Uysal, I., & Croft, W. B. (2011). User oriented tweet ranking: A filtering approach to

microblogs. Proceedings of the International Conference on Information and Knowledge Management (pp. 2261–2264). DOI: 10.1145/2063576.2063941.

Vergeer, M., Hermans, L., & Sams, S. (2010). The voter is only a tweet away: microblogging during the 2009 European Parliament election campaign in the Netherlands. Open SIUC Political Network Conference. Retrieved 29 May, 2016, from http://opensiuc.lib.siu.edu/pnconfs_2010/13.

Webberley, W. M., Allen, S. M., & Whitaker, R. M. (2016). Retweeting beyond expectation: Inferring interestingness in Twitter. Computer Communications, 73, 229–235.

Wu, P. F. (2013). In search of negativity bias: An empirical study of perceived helpfulness of online reviews. Psychology & Marketing, 30(11), 971-984.

Yan, R., Ogle, J. P., & Hyllegard, K. H. (2010). The impact of message appeal and message source on Gen Y consumers' attitudes and purchase intentions toward American Apparel. Journal of Marketing Communications, 16(4), 203-224.

Yang, Z., Guo, J., Cai, K., Tang, J., Li, J., Zhang, L. & Su, Z. (2010). Understanding retweeting behaviors in social networks. CIKM '10: Proceedings of the 19th ACM International Conference on Information and Knowledge Management (pp. 1633–1636). New York. DOI: 10.1145/1871437.1871691.

Hou, W., Huang, Y., and Zhang, K. (2015). Research of micro-blog diffusion effect based on analysis of retweet behaviour. Proceedings of the 2015 IEEE 14th International Conference on Cognitive Informatics and Cognitive Computing, pp. 255–261. doi: 10.1109/ICCI-CC.2015.7259394.

Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. Journal of Marketing, 74(2), 133-148.

Table 1 – Sender and message characteristics

	Appended by Twitter	Researched, added manually	Calculated from Twitter data	Appended by Brandwatch	Computer-based sentiment and content analysis	Manual sentiment and content analysis
Sender	Twitter handle Full name Twitter followers Twitter following Cumulative Twitter tweets Twitter verified	Party affiliation Gender of sender Age of sender Parliamentary cohort Election result Marginality of seat	Following ratio Campaign tweets per day Total campaign tweets Campaign retweets generated Sum of campaign replies Campaign retweets per tweet	Kred influence Kred outreach		
Message sentiment				Brandwatch: positive / negative / neutral	SPSS sentiment: positive / negative	Sentiment (manually-coded) : positive / negative / neutral
Message content					Election hashtags: #bbcqt #debate #voteconservative #votelabour #labourdoorstep #GE2015 SPSS concepts SPSS categories	Category of tweet topic / purpose (see Table 4)
Message structure	Twitter reply count Twitter reply to Number of times retweeted Twitter retweet of Thread entry type		Hashtag? yes / no Hashtag count URL link? yes / no Media link? yes / no Number of @ mentions	Tweet impact score Tweet impressions Reach		

Table 2 - Summary of Tweet content codes[1]

Content code	Tweet content	Example Tweets	Link to literature	Frequency	% total sample
Local	Mention of local constituency	Spring has sprung gloriously in Wycombe pic.twitter.com/fv3VnP6DRW (@SteveBakerHW)	Jackson & Lilleker 2011	459	37.9%
Campaign trail	Mention of some aspect of the day's campaigning on the doorstep	We spent this afternoon in Hawes Side for the #12Wards12Days challenge, discussing the cost of living with voters. Pic.twitter.com/3jVNPLdOPv (@GordonMarsden)	Amalgam of Graham et al.'s (2013) codes which include categories for updates from the campaign trail, campaign promotion and campaign action.	219	18.1%
Attack	Attacking some aspect of opposing party's policy or individual politicians from the other side	Labour campaign on NHS in Wales takes voters for idiots. Its Labour cuts that got us in this mess. (@GutoBebb)	No direct equivalent in existing schema. Graham et al. (2013) include a category for tweets which critique or argue and Sæbø (2011) has a category for tweets which feature dialogue with other politicians.	208	17.2%
Personal	Non-political tweet related to some other aspect of life	Congrats to @PtstudioWigan & all at #ptstudio on being crowned Gym of the Year. Still don't like #burpees tho!(@Y_FovargueMP)	Golbeck, Grimes, & Rogers (2010); Jackson & Lilleker (2011); Lawless (2012); Graham et al. (2013); Sæbø (2011)	149	12.3%
Thanks	Thanking people for help, volunteering, support etc.	Our unsung heros @IslingtonSouth: the stuffing team, the data team, the boards men. Thankyou! Pic.twitter.com/Gqp7mblZER (@EmilyThornberry)	Hemphill, Otterbacher, & Shapiro (2013), Graham et al. (2013)	141	11.6%
Position taking	Expressing an opinion on some aspect of own or opposition's policy	I believe Birmingham & Britain as a whole only succeed when working people succeed. #LabourManifesto sets out how labour.org.uk/page/-/Britain… (@RichardBurdenMP)	Lawless (2012), Graham et al. (2013)	129	10.6%

Achievement	Highlighting some personal achievement or achievement of the party.	So pleased to hear a passport is on its way to a constituents [sic] daughter – much of yesterday spent sorting the problem – huge thanks to HMPO (@carolinenokes) There are now 2million more people in work than in 2010. Largest increase in employment has been in the North West (@Andrew4Pendle)	New code but along similar lines to Sæbø's (2011) category for informing people of politicians' ongoing activities and Lawless's (2012) 'credit taking' category.	90	7.4%
Support for self	Passing on messages of support which MPs have received for themselves.	Glad to see my GP Dr Weir will be voting @UKLabour #labourdoorstep (@KarlTurnerMP)	New code	77	6.4%
Information provision	Providing information	8 questions you should ask about the internet of the things from the @guardian (& me☺ theguardian.com/media-network/… (@ChiOnwurah)	Hemphill, Otterbacher, & Shapiro (2013)	66	5.4%
Support for others	Expression of support for someone else from own side.	Good luck today @JustineGreening. #Putney - #VoteConservative (@S_Hammond)	New code	66	5.4%
Event	Mention of a campaign event that has attended or is planning to attend	On my way now to Shildon hustings #GE2015 (@HelenGoodmanMP)	New code	65	5.4%
Call to vote	Calling for people to vote	Riverside residents. Have you voted yet? Me+Michelle Corrigan, your Lab candidate – dedicated cllr – working for you #voteLabour #winNW15 (@LouiseEllman)	Hemphill, Otterbacher, & Shapiro (2013), Graham et al. (2013), Golbeck, Grimes, & Rogers (2010)	48	4%
Meeting people	Referring to individuals or groups of people that has met during campaign	Lovely to meet Sarah whos 92 and been a member of the Party since 1947. #VoteLabour pic.twitter.com/LFGRJFmtNU (@GordonMarsden)	New code	45	3.7%
Media response	In response to something that the MP has seen on TV or heard on the radio.	Boris show his true colours in last few minutes on #marr (@Debbie_abrahams)	New code	44	3.6%

Weather	Mentions of the weather	Wow. Reckon its hottest day of year already! Kids dropped off at mega maker holiday club with sun lotion on. Excitement levels very high.(@sbrine)	New code	44	3.6%
Call to action (not voting)	Requesting recipients to do something other than vote (e.g. register to vote, sign petition, volunteer, donate money, attend event)	Come and help me win in Cardiff Central! Sign up here: cardiffld.org.uk/volunteer (@JennyWillott)	Inspired by Hemphill, Otterbacher, & Shapiro (2013) who distinguish between requests for action that require the recipient to do something meaningful such as vote compared to those that require less significant activity, such as signing a petition or reading something. Graham et al. (2013) have a similar category – mobilizing and organizing – into which they put tweets that make a request for direct action of some kind such as signing a petition or joining a campaign team.	41	3.4%
Fear appeal	A particular form of negative tweet in which the MP explicitly warns of negative consequences if the other side wins.	Our economy is recovering. Don't let Labour wreck it! Vote Conservative today! (@chhcalling)	New code	39	3.2%
Other	Tweets which cannot be allocated to another category	The final countdown has begun in the L (@heidi_mp)	Almost all pre-existing coding categories have a catch-all 'other' category. Tweets coded as 'other' cannot also have any other code allocated to them.	30	2.5%
Charity	Mention of some charity that the MP supports or has worked with.	Such a fantastic charity Im currently visiting with @MarkHunter – training people to give first aid to children. Pic.twitter.com/5Tzp5FeqTx (@nick_clegg)	New code	20	1.7%

1. Content codes based on manual coding

Table 3 – CHAID models: Most predictive variables

Variable	Predicti-veness	Variable definition	Variable type*
CHAID Model 1: Message structure variables			
Media link	45.6%	Message includes at least one media link (calculated from Twitter data)	Structure
Hashtag	35.3%	Message includes at least one #hashtag (calculated from Twitter data)	Structure
Mention	8.8%	Message includes at least one @mention of another Twitter user (calculated from Twitter data)	Structure
URL	5.9%	Message includes at least one URL link (calculated from Twitter data)	Structure
Mention number	4.4%	Number of @mentions included in the message (calculated from Twitter data)	Structure
CHAID Model 2: Source variables			
Campaign tweets per day	43.0%	The average number of tweets that the MP sent per day of the campaign (calculated from Twitter data)	Source
Twitter following	24.4%	Number of Twitter profiles the MP is following (appended by Twitter)	Source
Cumulative Twitter Tweets	21.5%	Cumulative number of Tweets ever sent from Twitter profile (appended by Twitter)	Source
Following ratio	10.4%	Ratio of followers to following (appended by Twitter)	Source
CHAID Model 3: Message content and message sentiment variables			
Attack	40.0%	Attacking some aspect of opposing party's policy or individual politicians from the other side	Content
Support for others	17.8%	Expression of support for someone else from own side.	Content
Campaign trail	14.4%	Mention of some aspect of the day's campaigning on the doorstep	Content
Personal	9.4%	Non-political tweet related to some other aspect of life	Content
Position taking	8.3%	Expressing an opinion on some aspect of own or opposition's policy	Content
Fear appeal	7.8%	A particular form of negative tweet in which the MP explicitly warns of negative consequences if the other side wins.	Content
Thanks	3.3%	Thanking people for help, volunteering, support etc.	Content

Cont.

CHAID Model 4: Combined source, message structure, message content and message sentiment variables

Twitter followers	42.8%	Number that are following the MP's Twitter profile (appended by Twitter)	Source
Sentiment	12.8%	Message sentiment (manually coded)	Sentiment
Total campaign Tweets	8.4%	Number of Tweets posted during campaign (calculated from Twitter data)	Source
Percent majority before	5.6%	Percentage majority attained by the MP at previous general election (researched and added manually)	Source
Party affiliation	4%	MP political party affiliation (researched and added manually)	Source
Hashtag	4%	Message includes a #hashtag (calculated from Twitter data)	Structure
Media link	3.6%	Message includes a media link (calculated from Twitter data)	Structure
Support for others	3.2%	Message content includes expression of support for someone else from own party	Content
URL	2.8%	Message includes a URL link (calculated from Twitter data)	Structure
Personal	2.4%	Message content is non-political and related to some other aspect of life	Content

* Structure, content and sentiment are message-related variables. Source includes sender-related variables.

Table 4 - Retweeting of tweets: sentiment

	Message retweeted[2]		
Tweet message sentiment[1]	**No**	**Yes**	**Total messages retweeted**
Positive			
Number	217	225	442
Percentage	49.1%	50.9%	
Neutral			
Number	341	225	566
Percentage	60.1%	39.9%	
Negative			
Number	41	163	204
Percentage	20.2%	79.8%	
			1212

1. Message sentiment manually coded. 2. $\chi^2 = 96.73$, df =2, $p<.01$

Table 5 - Retweeting of Tweets: content[1]

Content of the Tweet	Retweeted? No		Yes		
Local	220	47.9%	239	52.1%	χ^2 =0.658, df 1, Sig .417
Campaign trail	86	39.3%	113	60.7%	χ^2 = 11.023, df 1, Sig .001 **
Attack	55	26.4%	153	73.6%	χ^2 = 53.046, df 1, Sig .000 **
Personal	103	69.1%	46	30.9%	χ^2 = 26.389, df 1, Sig .000 **
Thanks	77	54.6%	64	45.4%	χ^2 = 1.718, df 1, Sig .190
Position taking	50	38.8%	79	61.2%	χ^2 = 6.566, df 1, Sig 0.01 *
Achievement	43	47.8%	47	52.2%	χ^2 = 0.105, df 1, Sig .746
Support for self	40	51.9%	37	48.1%	χ^2 = 0.210, df 1, Sig .647
Information provision	30	45.5%	36	54.5%	χ^2 =0.440, df 1, Sig .507
Support for others	15	22.7%	51	77.3%	χ^2 = 19.900, df 1, Sig .000 **
Event	40	61.5%	25	38.5%	χ^2 = 4.034, df 1, Sig .045 *
Call to vote	21	43.8%	27	56.3%	χ^2 =0.643, df 1, Sig .422
Meeting people	23	51.1%	22	48.1%	χ^2 = 0.053, df 1, Sig .817
Media response	29	30.5%	66	69.5%	χ^2 = 14.742, df 1, Sig .000 **
Weather	22	50%	22	50%	χ^2 = 0.006, df 1, Sig .938
Call to action (not voting)	22	53.7%	19	46.3%	χ^2 = 0.305, df 1, Sig .581
Fear appeal	4	10.3%	35	89.7%	χ^2 = 24.729, df 1, Sig .000 **
Charity	9	45%	11	55%	χ^2 = 0.159, df 1, Sig .690

1. Content codes based on manual coding ** p <.01. * p < .05.

Figure 1 - Proposed antecedents of retweeting

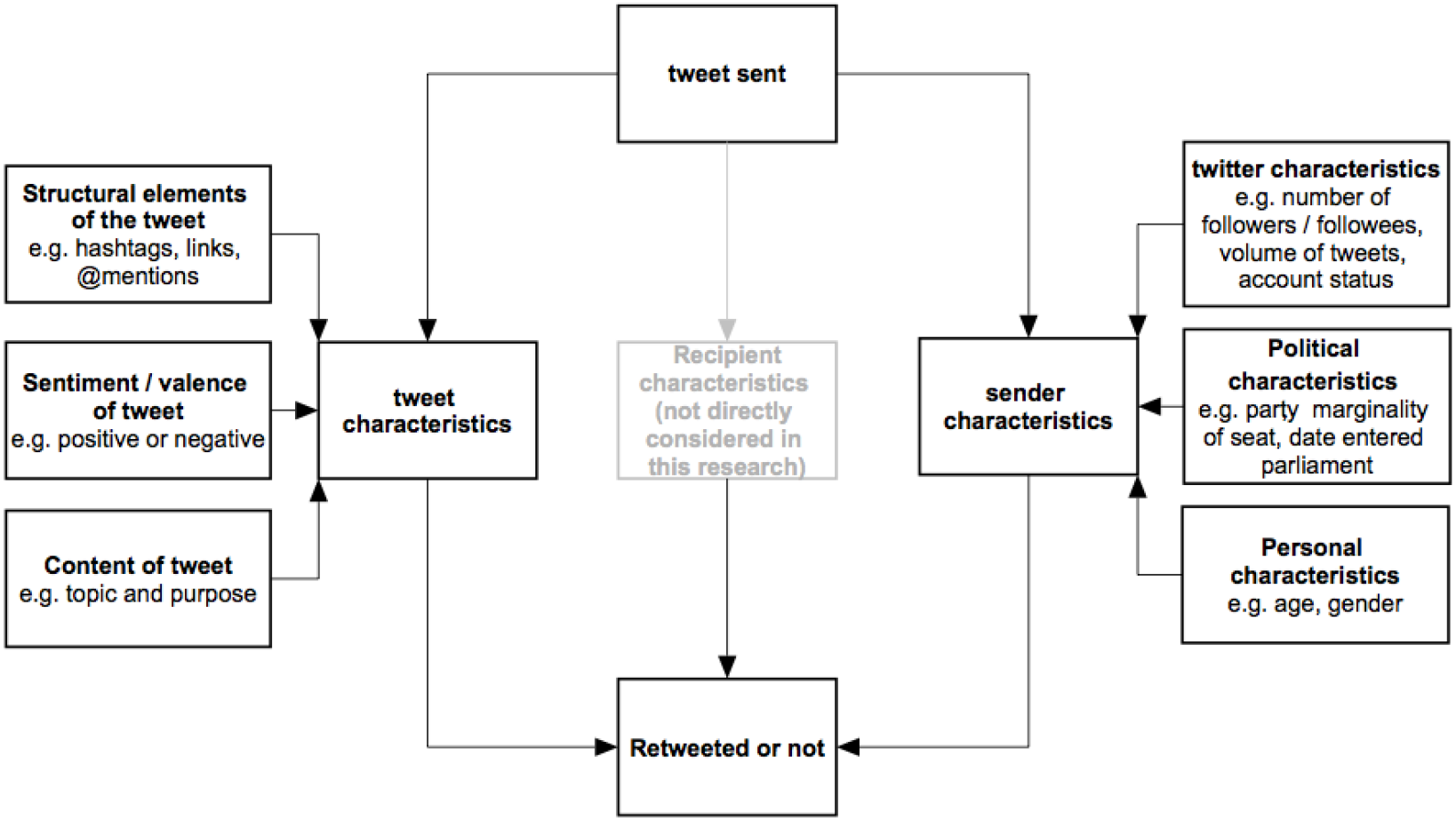
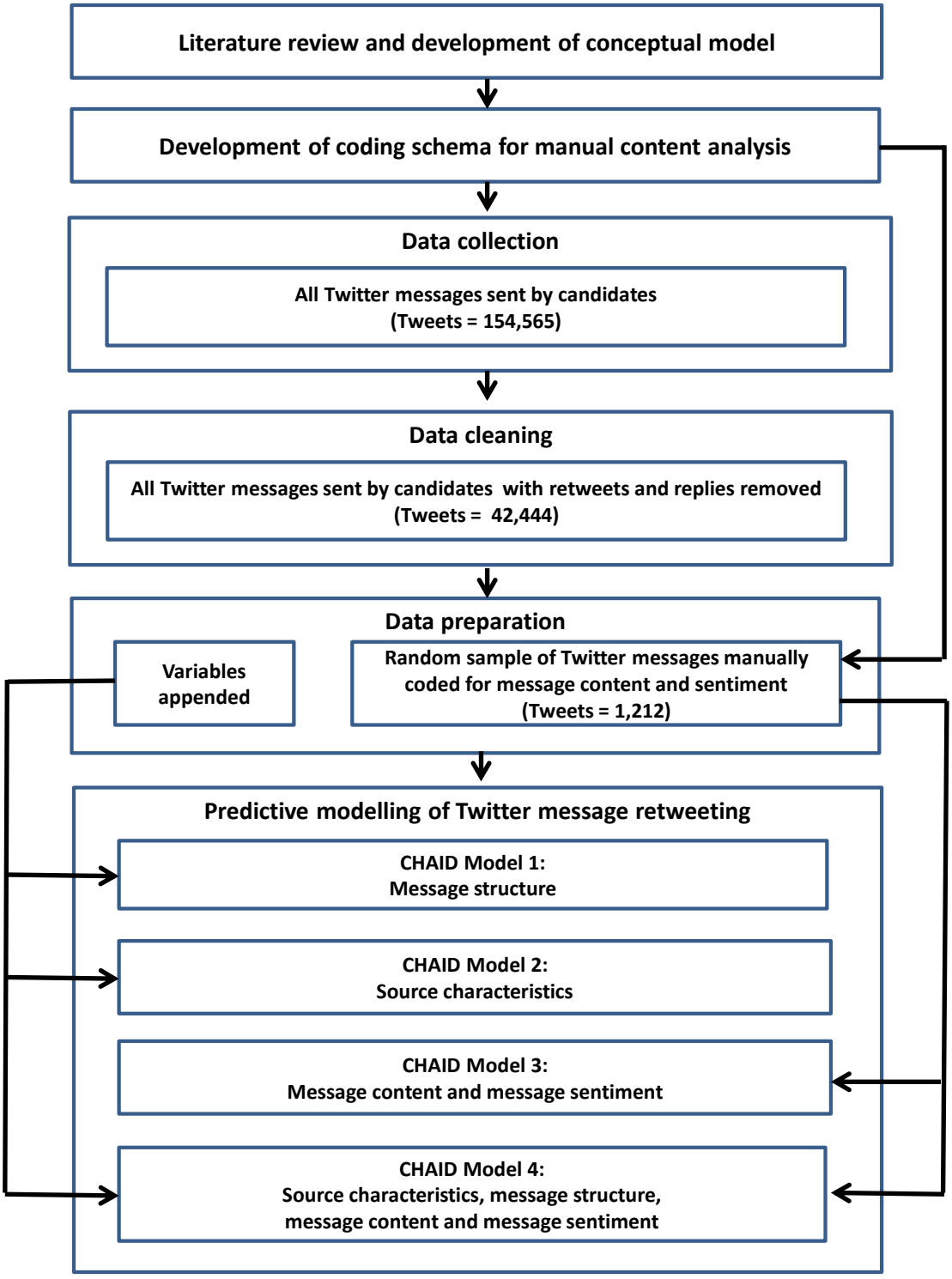


Figure 2: Methodological process for prediction of retweeting behavior



CHAID APPENDICES

ANTECEDENTS OF RETWEETING IN A (POLITICAL) MARKETING CONTEXT

Lorna Walker, Cranfield University

E: lorna.walker@cranfield.ac.uk

T.: +44(0)7939 084122

Paul R. Baines, Cranfield University (Corresponding author)

E: paul.baines@cranfield.ac.uk

T.: +44(0)1234 758041

Radu Dimitriu, Cranfield University

E: radu.dimitriu@cranfield.ac.uk

T.: +44(0)1234 751122

Emma K. Macdonald, Cranfield University

E: emma.macdonald@cranfield.ac.uk

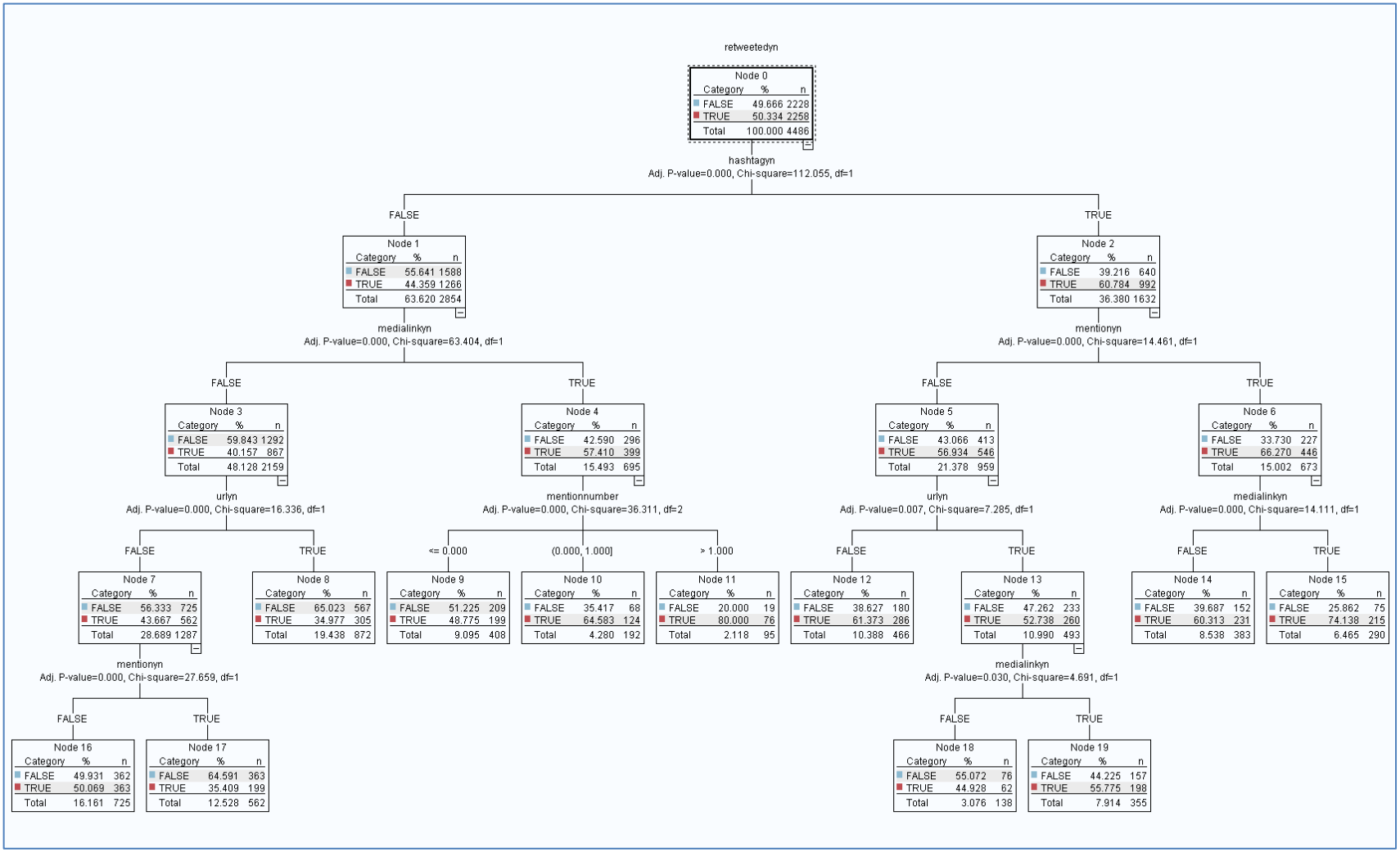
T.: +44(0)1234 751122

Appendix Table 1 – CHAID models

Variable	Predicti-veness	Variable definition	Variable type*
CHAID Model 1: Message structure variables (Appendix Figure 1)			
Media link	45.6%	Message includes at least one media link (calculated from Twitter data)	Structure
Hashtag	35.3%	Message includes at least one #hashtag (calculated from Twitter data)	Structure
Mention	8.8%	Message includes at least one @mention of another Twitter user (calculated from Twitter data)	Structure
URL	5.9%	Message includes at least one URL link (calculated from Twitter data)	Structure
Mention number	4.4%	Number of @mentions included in the message (calculated from Twitter data)	Structure
CHAID Model 2: Source variables (Appendix Figure 2, 2A, 2B, 2C, 2D, 2E, 2F)			
Campaign tweets per day	43.0%	The average number of tweets that the MP sent per day of the campaign (calculated from Twitter data)	Source
Twitter following	24.4%	Number of Twitter profiles the MP is following (appended by Twitter)	Source
Cumulative Twitter Tweets	21.5%	Cumulative number of Tweets ever sent from Twitter profile (appended by Twitter)	Source
Following ratio	10.4%	Ratio of followers to following (appended by Twitter)	Source
CHAID Model 3: Message content and message sentiment variables (Appendix Figure 3)			
Attack	40.0%	Attacking some aspect of opposing party's policy or individual politicians from the other side	Content
Support for others	17.8%	Expression of support for someone else from own side.	Content
Campaign trail	14.4%	Mention of some aspect of the day's campaigning on the doorstep	Content
Personal	9.4%	Non-political tweet related to some other aspect of life	Content
Position taking	8.3%	Expressing an opinion on some aspect of own or opposition's policy	Content
Fear appeal	7.8%	A particular form of negative tweet in which the MP explicitly warns of negative consequences if the other side wins.	Content
Thanks	3.3%	Thanking people for help, volunteering, support etc.	Content

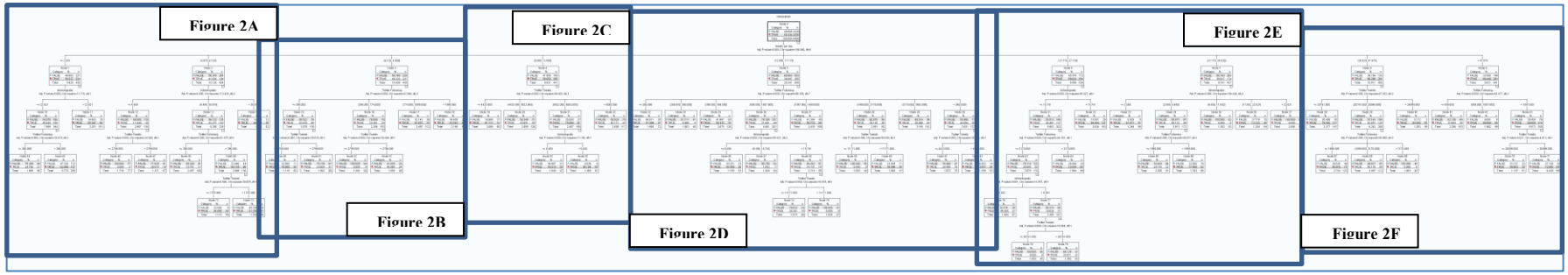
* Structure, content and sentiment are message-related variables. Source includes sender-related variables

Appendix Figure 1 - CHAID Model 1: Message structure variables only*

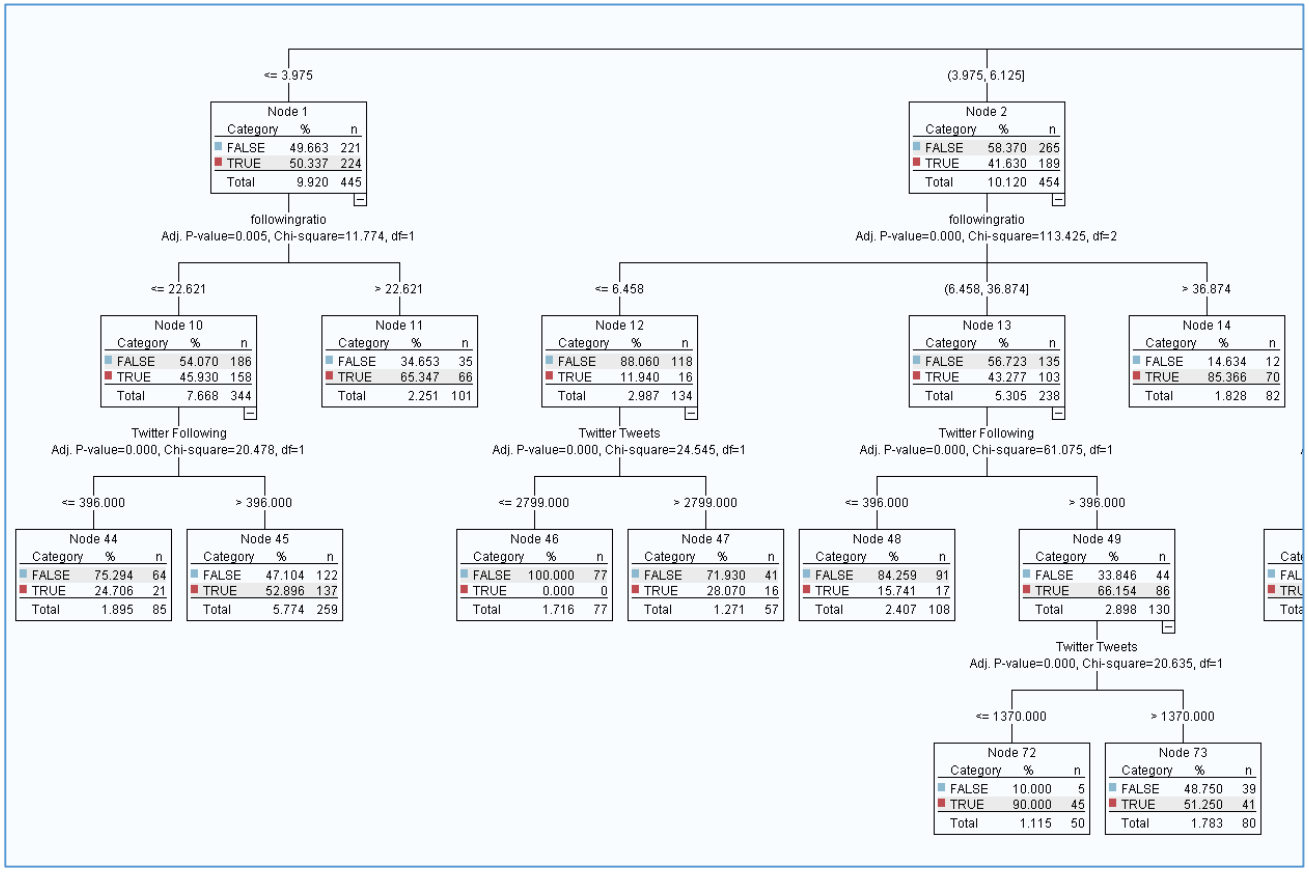


* Variables in Model 1: ***hashtagyn*** (does the tweet contain a hashtag? yes or no); ***medialinkyn*** (does the tweet contain a link to an image or video? yes or no); ***urlyn*** (does the tweet contain a link of any kind? yes or no); ***mentionnumber*** (how many mentions are there in the tweet?); ***mentionyn*** (does the tweet contain a mention? yes or no)

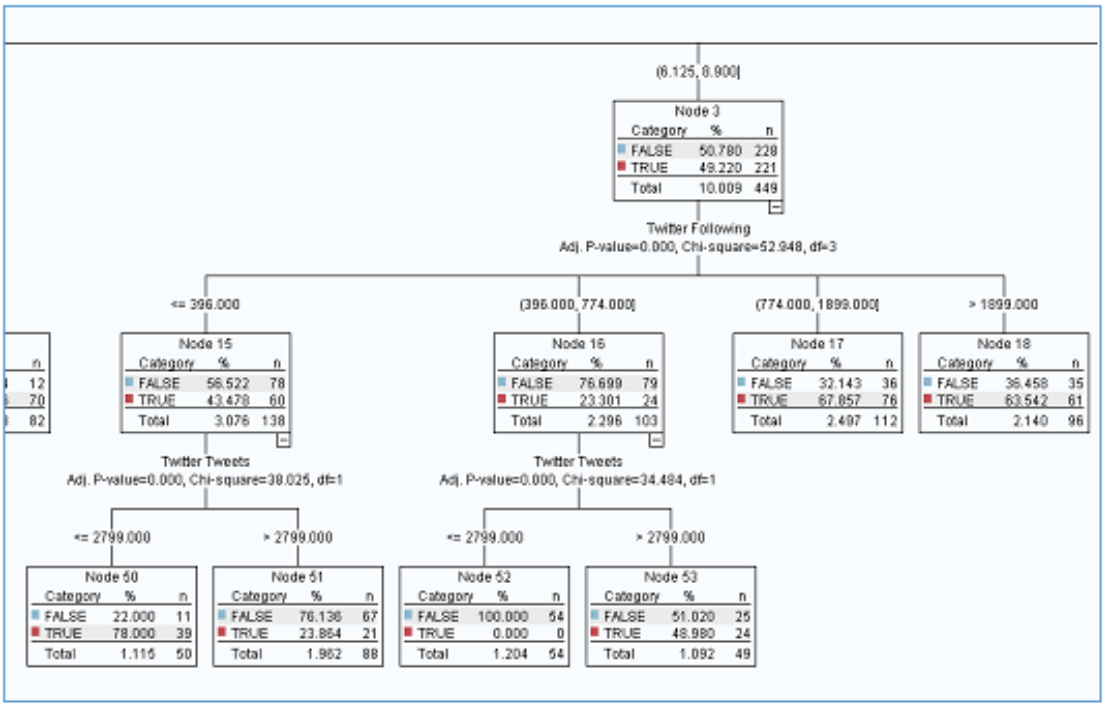
Appendix Figure 2 - CHAID Model 2: Source variables only*



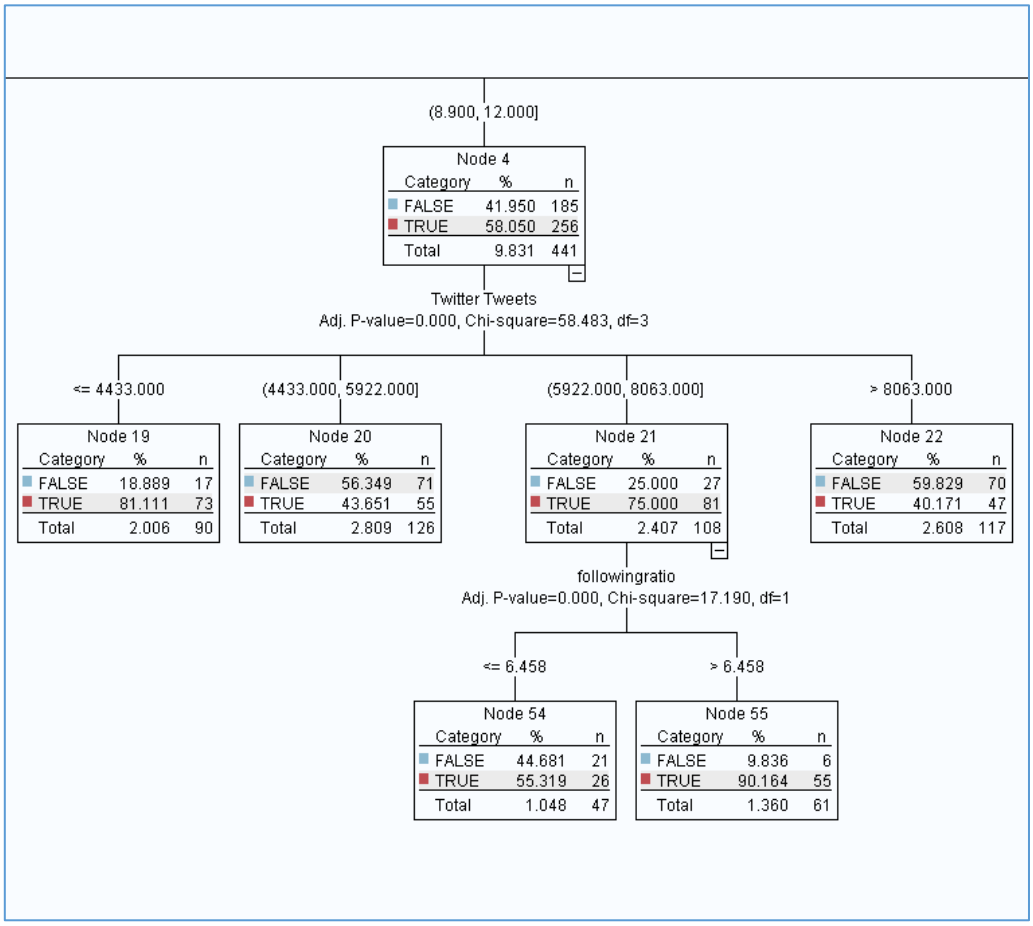
Appendix Figure 2A - CHAID Model 2: Source variables only (Part A)*



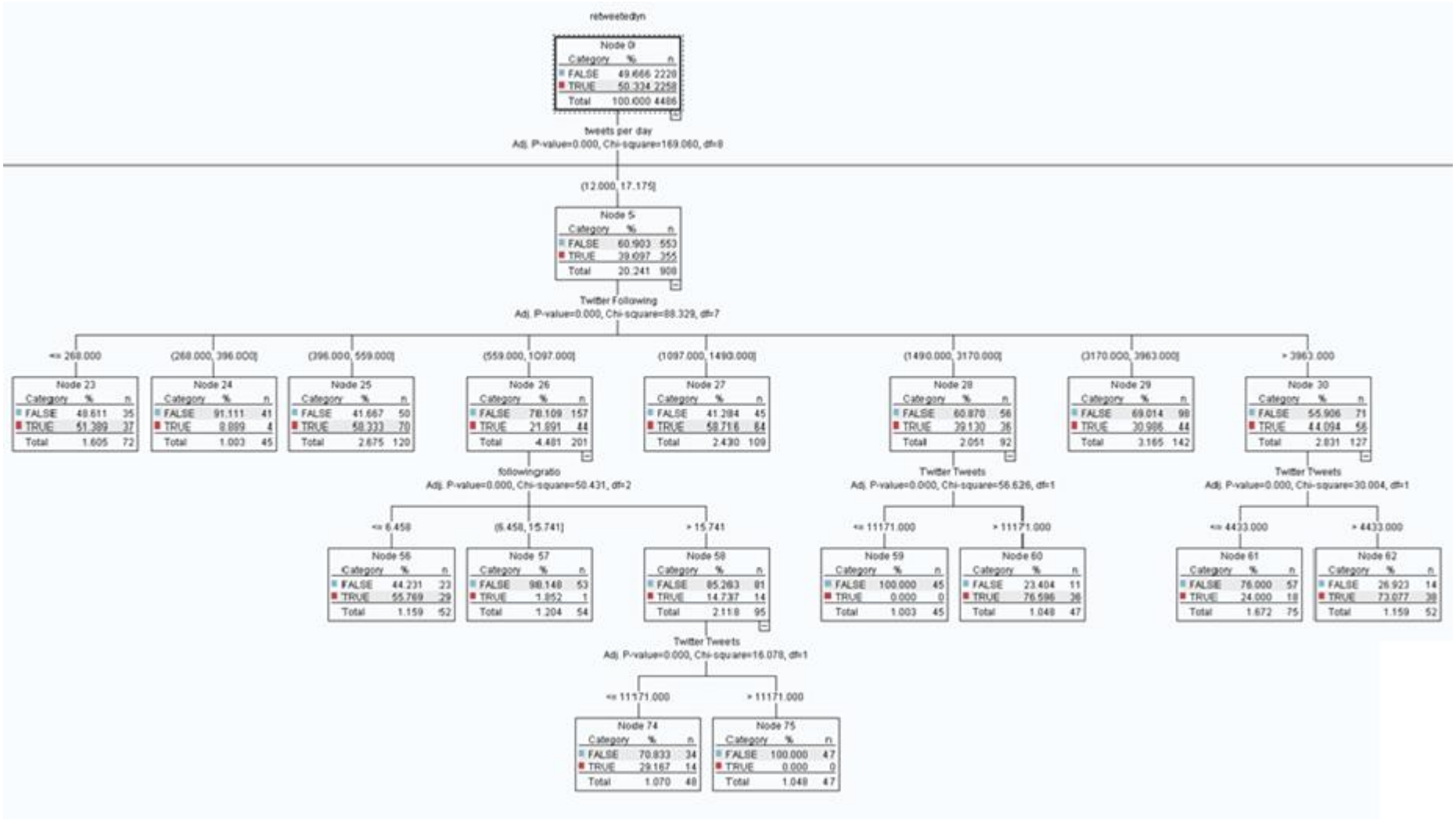
Appendix Figure 2B - CHAID Model 2: Source variables only (Part B)*



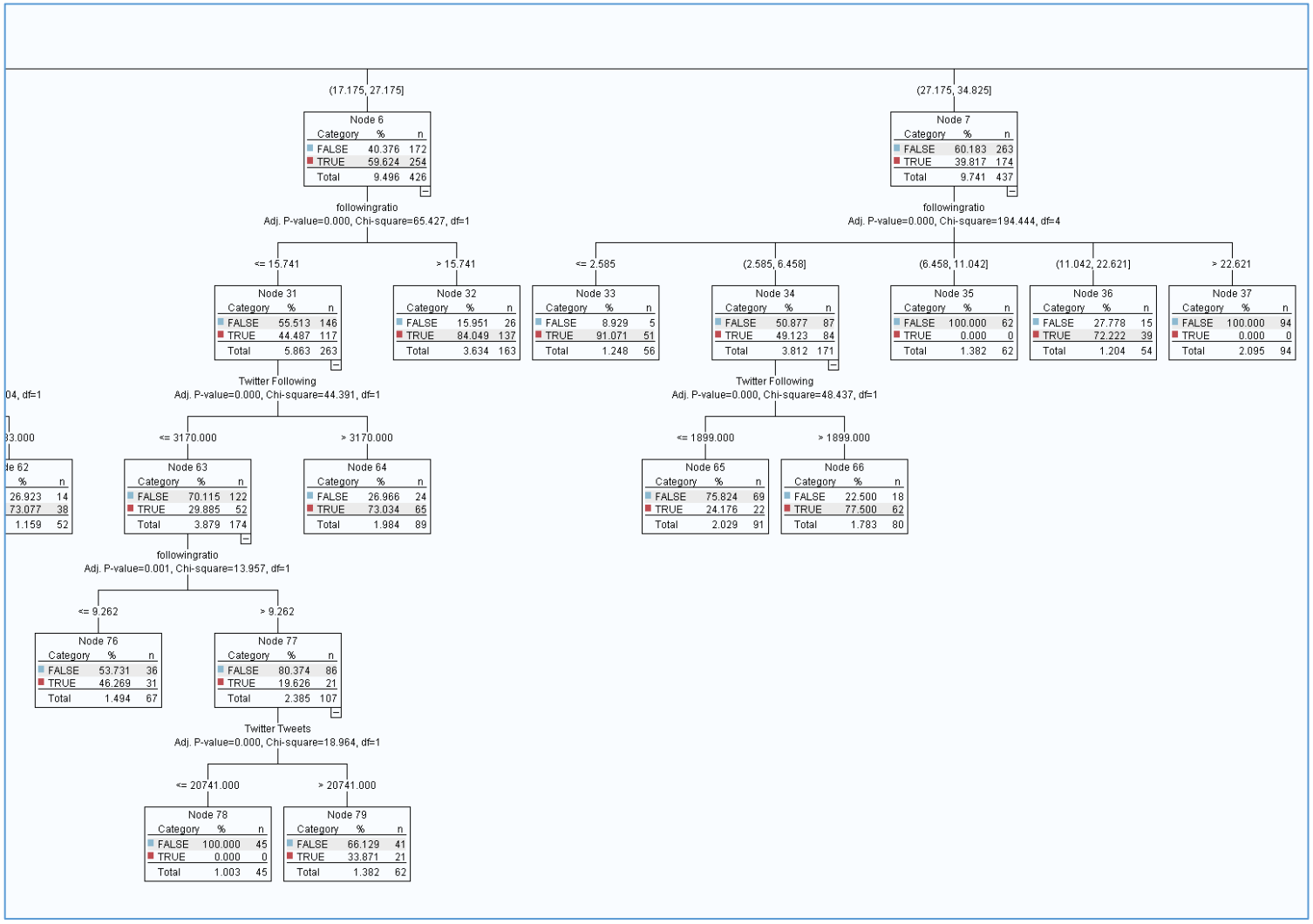
Appendix Figure 2C - CHAID Model 2: Source variables only (Part C)*



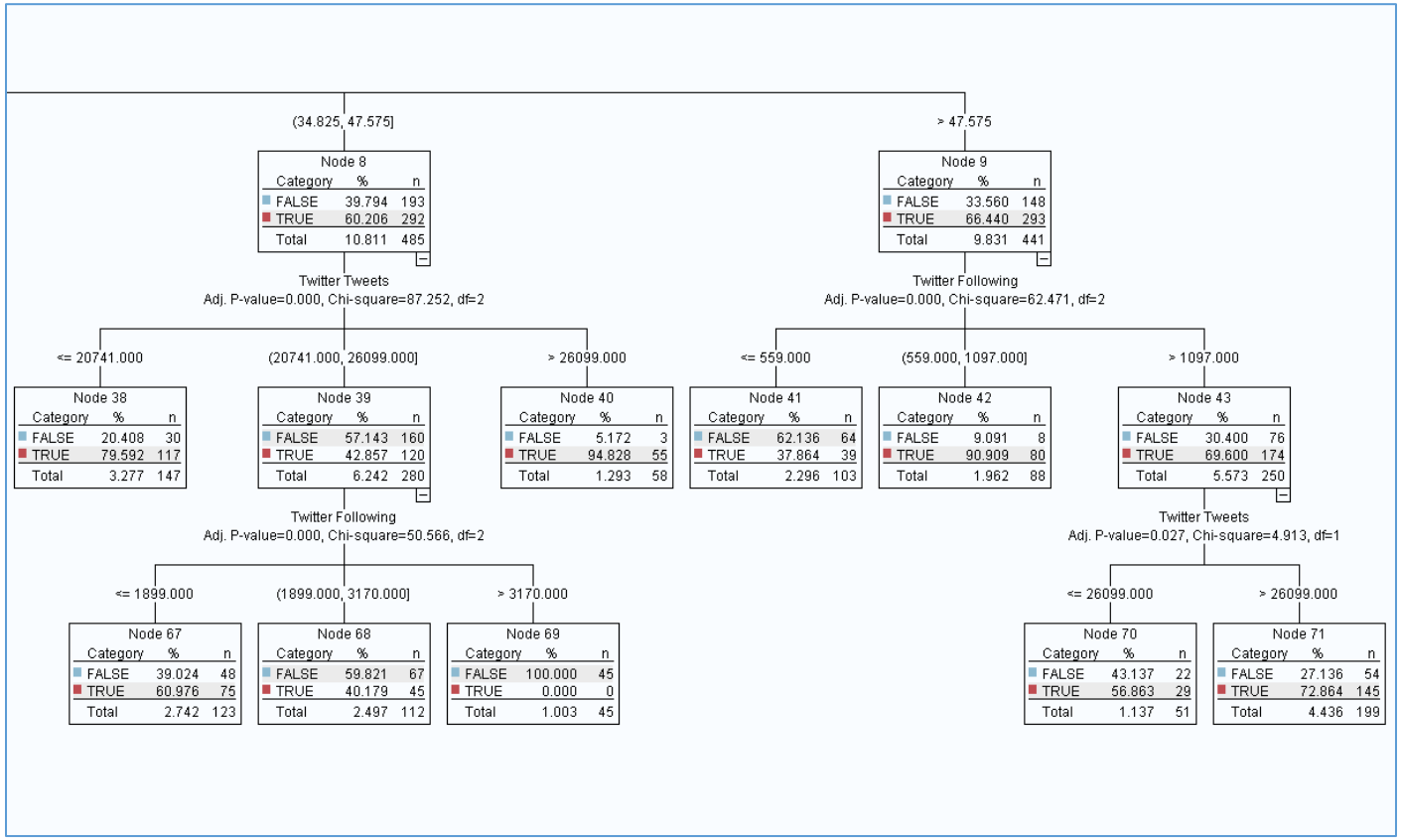
Appendix Figure 2D - CHAID Model 2: Source variables only (Part D)*



Appendix Figure 2E - CHAID Model 2: Source variables only (Part E)*

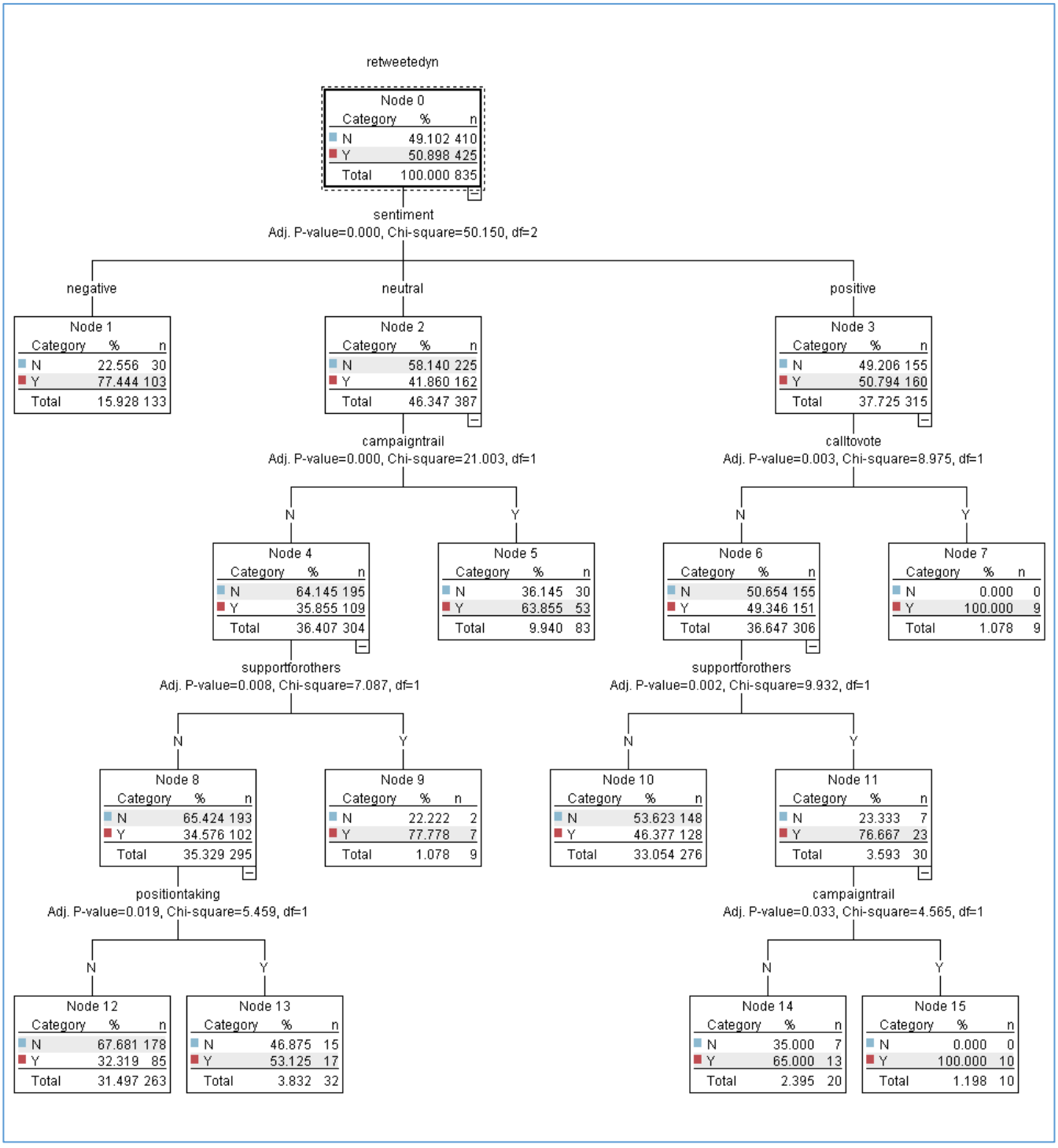


Appendix Figure 2F - CHAID Model 2: Source variables only (Part F)*



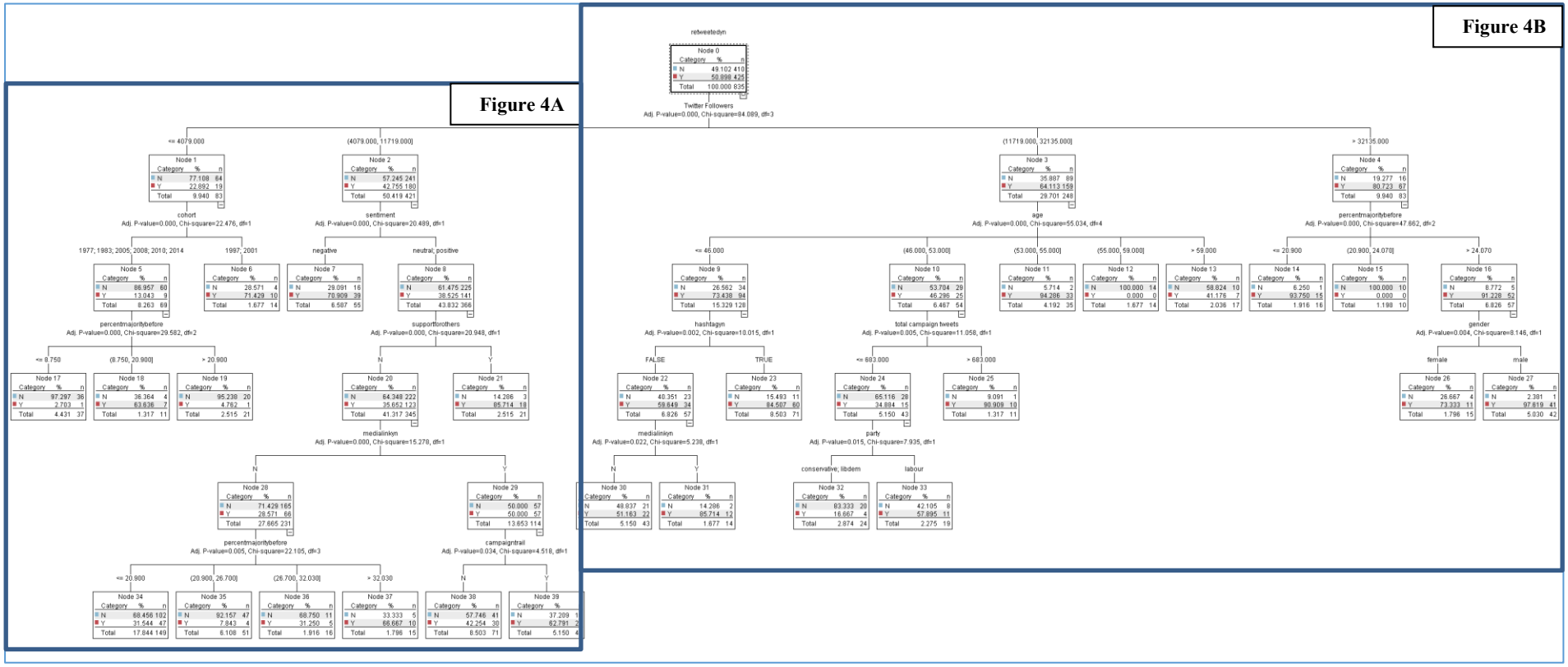
*Variables in Model 2: ***Twitter following*** (number of people the MP follows); ***Followingratio*** (number of followers divided by number of followees); ***Tweets per day*** (the number of tweets that the MP sent per day of the campaign); ***Twitter tweets*** (total number of tweets sent by the MP in their time on Twitter)

Appendix Figure 3 - CHAID Model 3: content and sentiment variables*

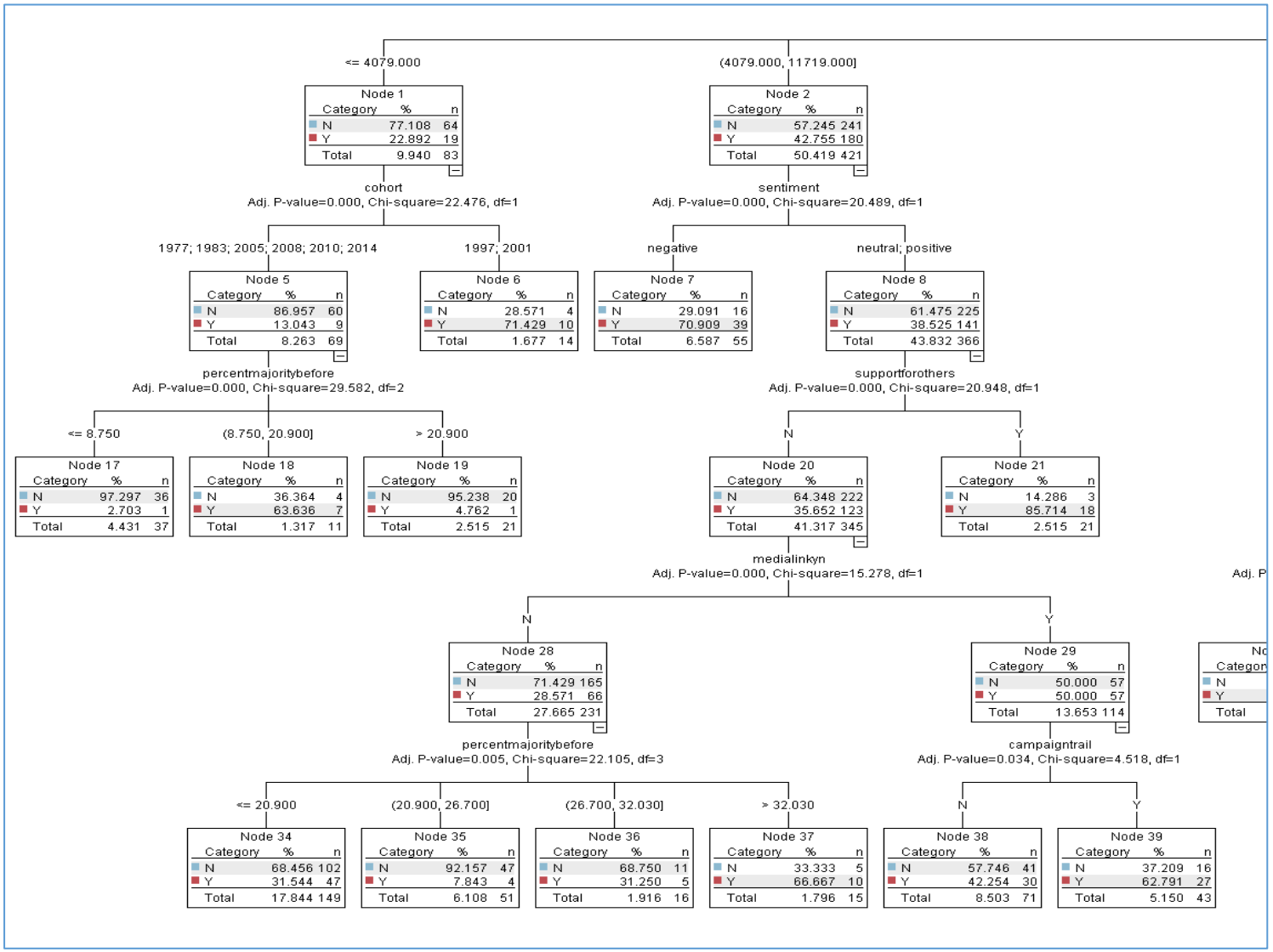


* Variables in Model 3: ***Sentiment*** (manual sentiment code: positive, negative or neutral); ***Campaigntrail*** (does the tweet include a mention of the campaign trail? yes or no); ***Calltovote*** (does the tweet include a call to vote? yes or no); ***Supportforothers*** (does the tweet express support for others? yes or no); ***Positiontaking*** (does the tweet suggest taking a position on an issue? yes or no)

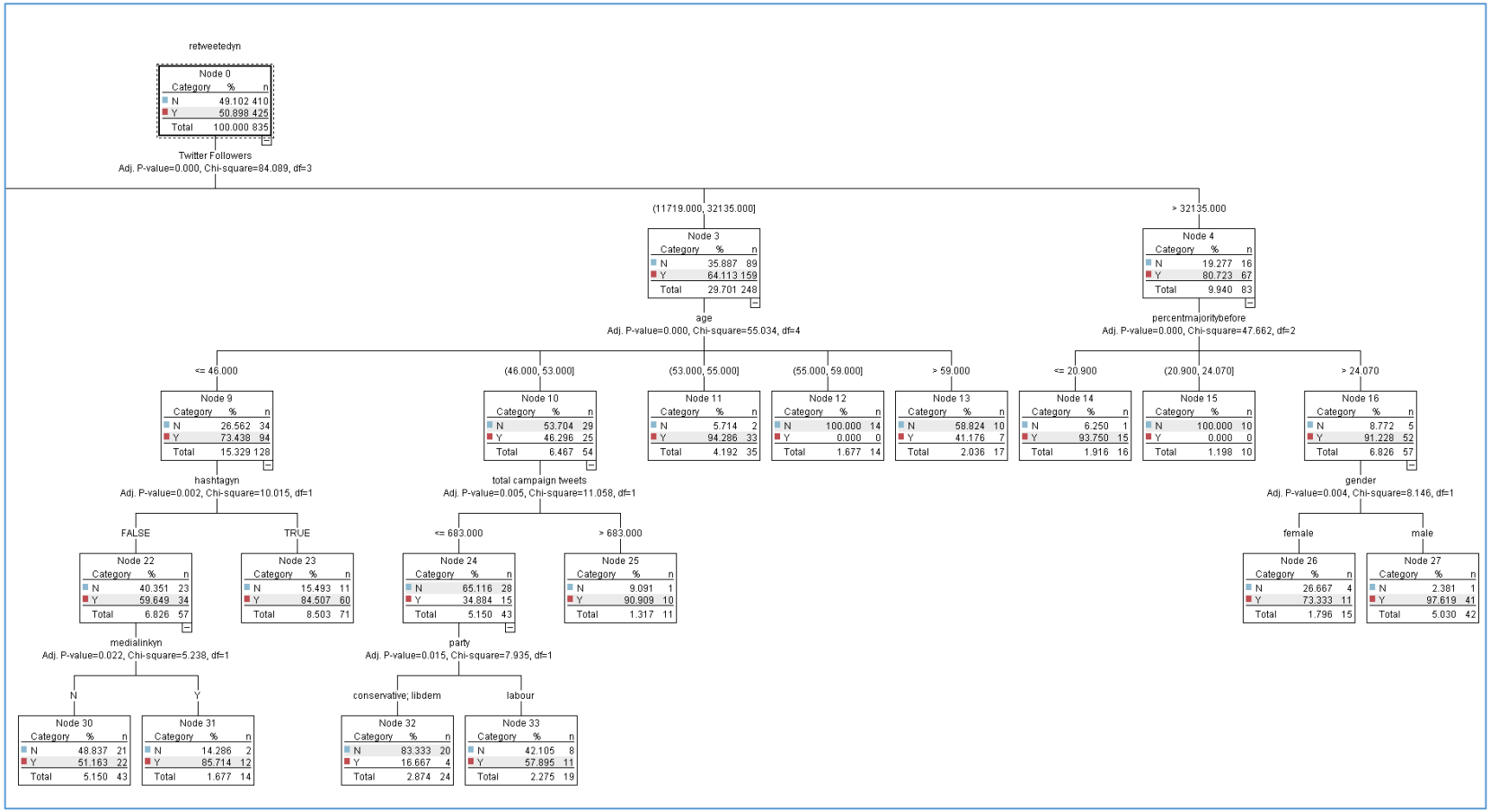
Appendix Figure 4 - CHAID Model 4: Combined source, message structure, message content and message sentiment variables*



Appendix Figure 4A - CHAID Model 4: Combined (part A)*



Appendix Figure 4B - CHAID Model 4: Combined (Part B)*



*Variables in Model 4: ***Twitter Followers*** (number of Twitter followers of MP***) Cohort*** (year the MP entered parliament); ***Sentiment*** (manual sentiment code: positive, negative or neutral); ***Percentmajoritybefore*** (the MP's % majority before the election); ***Supportforothers*** (does the tweet express support for others? yes or no); ***Medialinkyn*** (does the tweet include a link to an image or video? yes or no); ***Campaigntrail*** (does the tweet include a mention of the campaign trail? yes or no); ***Age*** (the MP's age); ***Hashtagyn*** (does the tweet include at least one #hashtag? yes or no); ***Total campaign tweets*** (total number of tweets the MP sent during the campaign); ***Gender*** (MP's gender).