Labels and sentiment in social media: On the role of perceived agency in online discussions of the refugee crisis

Ju-Sung Lee
Erasmus University Rotterdam
P.O. Box 1738
Rotterdam, The Netherlands 3000 DR
lee@eshcc.eur.nl

Adina Nerghes
Vrije Universiteit Amsterdam
De Boelelaan 1081
Amsterdam, The Netherlands 1081 HV
adina.nerghes@vu.nl

ABSTRACT

By focusing on the recent events in the Middle East, that have pushed many to flee and seek refuge in neighboring countries or in Europe, we investigate dynamics of label use in social media, the emergent patterns of labeling that can cause further disaffection and tension, and the sentiments associated with the different labels. To achieve this, we examine key labels pertaining to the refugee/migrant crisis and their usage in the user comment thread of a highly viewed and informational video of the crisis on YouTube. The use of labels indicate that migration issues are being framed not only through labels characterizing the crisis but also by their describing the individuals themselves. The sentiments associated with these labels depart from what one would normally expect; in particular, negative sentiment is attached to labels that would otherwise be deemed neutral or positive. Interestingly, both positive and negative labels exhibit increased negativity across time. Using topic modeling and sentiment analysis jointly, we discover that the latent topics of the most positive comments show more overlap than those topics of the most negative comments, which are more focused and partitioned. In terms of sentiment, we find that labels indicating some degree of perceived agency or opportunity, such as 'migrant' or 'immigrant', are embedded in less sympathetic comments than those labels indicating a need to escape war-torn regions or persecution (e.g., asylum seeker or refugee). Our study offers valuable insights into the direction of public sentiment and the nature of discussions surrounding this significant societal event, as well as the nature of online opinion sharing.

CCS CONCEPTS

• **Information systems** → **Social networking sites;** *Sentiment analysis;* • **Computing methodologies** → *Topic modeling;* • **General and reference** → *Empirical studies;* • **Applied computing** → *Sociology;*

KEYWORDS

Social media; Sentiment analysis; Topic modeling; Labels and frames; European refugee crisis.

#SMSociety'17, Toronto, ON, Canada

978-1-4503-4847-8/17/07…$15.00
DOI: 10.1145/3097286.3097300

1. INTRODUCTION

In recent years, the precarious and unstable situation in the Middle East has pushed many to flee their countries and seek refuge in neighboring countries or in Europe. In April 2015, when five boats sank in the Mediterranean Sea, killing more than 1,200 people, the phrases 'European migrant crisis' and 'European refugee crisis' started being widely used by media and politicians alike. By using such labels to describe these events, the influx of displaced people into Europe has been framed in specific ways. Such labels, serve as frames that alter perceptions and perhaps even influence behaviours. For instance, while the use of 'refugee' portrays people fleeing armed conflict or persecution, 'migrant' describes people making a conscious choice to leave their country to seek a better life elsewhere. These dichotomized characterizations can have serious consequences for the lives and safety of asylum seekers; they can undermine public support, steer public opinion, and frame the debate on how the world should react to this crisis.

The use of labels has the potential to shape the range of possibilities for understanding what the story is on migration, and the way we perceive migrants and refugees. Negatively labeling and framing of refugees and migrants across Europe – and beyond – may lead to serious problems at the level of the host societies, where perceptions are significantly influenced. For example, as shown by the report issued by the European Commission (2006), public perception of migration tends to be increasingly negative throughout Europe. Thus, labels and frames provide indications of the ways in which displaced people are received and perceived worldwide.

While the use of powerful labels to characterize events reported in the media have become the norm, the recent influx of displaced people into European countries has also lead to a number of 'unfortunate' statements from politicians. A few examples come from British Prime Minister David Cameron who stated that "swarms of people [are] coming across the Mediterranean"[52] and Home Secretary David Blunkett who referred to child asylum seekers as 'swamping' some British schools [14]. French politician Marine Le Pen referred to the wave of refugees as "migrant anarchy," [16] while Luxembourg's Foreign Minister Jean Asselborn declared that "we are heading into anarchy" when discussing the refugee crisis [2].

The labels used by the media, politicians, and even online information sources, clearly indicate that migration issues are being framed by labeling the event but also by labeling the individuals themselves. Employing certain labels, keywords, or stock phases (e.g., refugee, refugee crisis, migrant, migrant crisis, immigrant,

immigrant crisis, Syrian, Syrian migrant, Syrian refugee, asylum seeker etc.) in communication contexts may affect receivers by emphasizing different frames for evaluation of the same issue or event [e.g., 10, 13, 15, 19, 49].

Through framing, certain features of a story are selected while others are excluded [29], and frames may shape people's interpretation of that story by making certain perspectives more salient [23, 29, 43]. Drawing from the work of Goffman [19], we understand that frames elicit, as well as constrain, the interpretative activities of audiences [43]. Entman [15] defines framing as a way "...to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation." By highlighting certain characteristics of an issue and hiding others, framing reflects the emphasis of the author.

While recognizing the importance of frames and frame analysis, we focus on the use of labels in this article. The close relationship between labeling and framing is implicitly acknowledged by employing the term "framing labels" [33]. In other words, we consider labels as the building blocks in the creation of frames, and we postulate that the selection and use of labels is a crucial and important instrument in the process of framing particular events and individuals.

Whether rooted in cognitive psychology [e.g., 31, 57], or the social sciences [e.g., 13], most studies focus on the analysis of frames and labels used in news media or public discourse, and their varying effects on people's choices or attitudes. Many examples of studies investigating media use of frames and labels on migration issues can be found [e.g., 24, 28, 47, 48], as well as examples of studies investigating their effects [e.g., 1, 6, 7]. However, in this study, we take a different approach by uncovering the use of labels in social media and the sentiment surrounding these labels. Thus, we do not focus on the labels employed by the mass media or European political figures. Rather, we draw a parallel between the various labels attributed to this recent crisis by laypeople in social media and the various sentiments associated with each label.

The battle over the words used to describe migrants, or the "struggle over framing" [17], does not take place only in mass media or public discourses. Rhetorical framing labels have become an integral part of social media and the online world. An example is the Wikipedia entry on the topic of the "European migrant crisis," which begins by stating "The European migrant crisis or European refugee crisis began in 2015..." [58]. The use of these two (distinct) labels joined by 'or' seems to imply they are equivalent or synonymous. However, as proposed earlier, the terms 'migrant' and 'refugee' denote very distinct characteristics of the individuals labeled as such.

With this study, we aim to uncover the dynamics of label use in social media surrounding the recent influx of displaced Middle-Eastern individuals, the emergent patterns of labeling that can cause further disaffection and conflicts, and the sentiment associated with the different labels. For our data collection, we focus on YouTube, one of most frequently visited Internet sites that stimulates social interactions through user-generated content, such as comments and responses to comments.

1.1 User-Generated Content

In recent years, the way in which people use the Internet has been evolving, with a remarkable shift towards increased user participation in creating and uploading content (photos, videos, audio and textual information), sharing and recommending content, and posting comments and ratings on the user-generated content, as well as on the online resources relevant to that content [18, 25]. While the exact role of new media is still being debated, there is no denying that for many people social media has become a source of information, important influencer of emotions and a way to organize activities and make decisions [54].

YouTube, in particular, has become one of most frequently visited Internet sites, with more than 100 million videos viewed daily. The popularity of YouTube can be attributed in part to the ability of individuals to both retrieve and post information. In this sense, YouTube can be considered a hybrid form of communication because it serves as a mass media form of communication, but also as an interpersonal form of communication. As a form of mass media communication, YouTube allows users to upload videos that others can view in a one-to-many approach. On the other hand, YouTube stimulates social interactions, allowing users to view and post videos and written responses in a one-to-one approach [37].

2. PREVIOUS RESEARCH

Recent efforts have increasingly focused on studying YouTube user behaviour by measuring video popularity and video content through quantitative analysis methods. For example, Paolillo [44] analyzed user profiles, including friending, subscription, favorite-ing and commenting, and identified that certain types of content were cultivated by users from particular social groups with shared characteristics. Similarly, Canali et al. [8] assessed the strength of links between users in terms of in-degrees and out-degrees, and they found that certain users had a significantly higher proportion of fans in relation to invited friends, and termed these people 'hub-users'. Chatzopoulou et al. [9] analyzed over 37 million videos, investigating properties like view counts, number of comments, ratings given and number of times a video is tagged as a 'favorite', in order to uncover the best indicators of video popularity. Their results suggest that favorite-ing, commenting or rating was a stronger indicator of popularity than simply viewing a video, because it requires more effort to log in to express a reaction. Kousha et al. [34] provide a more comprehensive review of quantitative studies investigating YouTube videos in a multitude of domains including marketing, medicine and management.

In recent years, qualitative studies investigating YouTube data have also started to emerge. For example, Lange [36] analyzes video sharing behaviour, which reveals complex interaction patterns between YouTube users are becoming available. Another qualitative study of YouTube content was conducted by Kousha et al. [34], and it examines the type of YouTube videos cited in academic publications. However, most of the studies summarized above focused on the video type, content, or video statistics (e.g., number of views, likes, dislikes etc.), rather than on the content of user-generated comments. YouTube comments have, thus far, been comparatively understudied in relation to other aspects of the site. The large number of comments, lack of structured organization, and the variable quality in terms of spelling, grammar and expression, have presented considerable difficulties for such studies. However, as Siersdorfer et al. [53] argue, "the amount of community feedback in YouTube results in large annotated comment sets which can help to average out noise in various forms and, thus, reflects to a certain degree the "democratic view of a community." Still, research has found that YouTube comments appear to reflect real-life communication behaviour [51]. Rojas et al. [50] use the term "media dialogue" to describe the ability of media messages to serve as a springboard for discussion.

YouTube commentary can sometimes serve as a lens for public opinion on issue importance, or even as a source for user mobilization, learning, and opinion-formation [32]. For example, participatory dynamics on YouTube comments surrounding 'climategate' have been investigated through a qualitative multideterminant framework [45]. Jones and Schieffelin [30] used comments to investigate language use associated with particular genres of video through qualitative research methods and a small number of videos.

The capability of YouTube to stimulate social interactions through the user comments section makes it a valuable site for investigating the use of labels in response to the issues arising from the recent influx of individuals from war-torn and/or economically challenged countries. Social media platforms, such as YouTube, exist as the continual co-creation of millions of participants and they depend on individual and collective participation and creation. Social media responses to societal events – in the form of comments for instance – are used for self-expression (positive or negative), providing emotional support, reminiscence, grieving and advice, as well as direct comments on the video itself [39]. Such responses are often characterized by the relative anonymity of personal expression and may lead to expressions of empowered and uninhibited public opinion. YouTube users believe that sensitive or uncomfortable topics are more easily discussed in online settings like YouTube [35]. Hence, YouTube comments have the potential to expose the ways in which labels are used and the affective content surrounding them. While labels themselves can be positive or negative, we explore the sentiment of the content surrounding the various labels employed by YouTube users, how the sentiment evolves over time, and also the emergent topics in these comments and how the most prominent labels manifest in these topics. By investigating the use of labels in social media commentary, discussion of alternative perspectives may be uncovered and understood [41].

3. DATA AND METHODS

For our analysis, we selected a highly popular YouTube video (which has been viewed over 10 million times) on the topic of interest, namely, "The European Refugee Crisis and Syria Explained"[1]. This is a six minute video published on September 17, 2015 that offers an objectively sympathetic perspective to the crisis, through animation, voice-over, and a musical score. 46,313 publicly accessible comments posted to this video were collected on October 22, 2016 using Netvizz YouTube Data Tool [46], and they constitute our text corpus. These comments represent both top-level comments ($n = 16719$) and their replies ($n = 29594$) posted between the video's upload date up to October 10, 2016 (over one year). These data were cleaned prior to analysis, specifically, noise-words, punctuation, and numbers were removed. Additionally, all words were lowercased and stemmed (i.e., words were reduced to their morphemes, such as plurals converted to singular forms).

Table 1. Broader Codes

Code	Labels
Refugee	refugee, refugee crisis
Syrian	syrian, syrian migrant, syrian refugee
Migrant	migrant, migrant crisis
Immigrant	immigrant, immigrant crisis
Threat	jihadist, terrorist, criminal

The analysis of this study focuses on those comments containing labels that describe various aspects of the refugee/migrant crisis: refugee, refugee crisis, migrant, migrant crisis, immigrant, immigrant, crisis, syrian, syrian migrant, syrian refugee, asylum seeker, jihadist, terrorist, criminal, scum, muslim, islam, rapefugee. We refer to both the unigrams and bigrams (i.e., two word phrases) as 'labels' and each of these labels are exclusively coded. For example, a comment containing 'refugee crisis' will not be coded has having 'refugee'. In selecting these labels, a unique concept (unigrams and bigrams) list was generated and parsed by both authors to identify those labels directly relevant to the topic under investigation. Furthermore, to capture the broader uses of the identified labels, we code umbrella indicator variables for groups of labels shown in Table 1, and we will subsequently refer to these as 'codes'.

We perform sentiment analysis and topic modeling on the corpus of YouTube comments.

3.1 Sentiment Analysis

For sentiment analysis, we employ Thelwall's SentiStrength, which provides scores of two dimensions of sentiment (positivity and negativity) per emotional term and phrases within the comment[2], while ignoring the sentiment of the key terms [55]. We are interested in *how* those labels are discussed, so their accompanying sentiments are discounted from the sentiment scores; that is, the key labels were removed from the corpus prior to sentiment analysis in order to assess the sentiment of each comment absent any label.

Sentiment scores from SentiStrength range from 0 to 4 for capturing the extent of the Positive and Negative sentiment dimensions in a segment of text (i.e., a YouTube comment)[3]. Additionally, we create a new variable for capturing sentiment on a single dimension ranging from -4 to $+4$; for this, the negative sentiment score (0-4) was

[1]The URL for the video is https://www.youtube.com/watch?v=RvOnXh3NN9w

[2]In addition to the lexicon based sentiment identification, SentiStrength also assigns sentiment to emoticons based on a list with human-assigned sentiment scores [55].

[3]The initial SentiStrength scores of -1 to -5 for negativity and $+1$ to $+5$ for positivity were recoded into the 0-4 range in which higher numbers indicate more intense sentiment, and -1 and $+1$ represent neutrality in the software.

multiplied by -1. The calculation for computing the unidimensional Sentiment score is:

$$\text{Sentiment} = \text{Positivity} + (-1)\text{Negativity}.$$

Also, we capture intensity of sentiment considering both the extent of positivity and negativity; this new variable is calculated as the Euclidean distance of the two dimensions score to neutrality (i.e., (0,0)):

$$\text{Intensity} = \sqrt{\text{Positivity}^2 + \text{Negativity}^2}.$$

3.2 Topic Modeling

For topic modeling, we employed latent Dirichlet allocation (LDA) [4], a three-level hierarchical Bayesian model, as implemented in the MALLET [40] library used by the text analysis software, ConText [12]. Topic models are a class of automated text analysis tools that seek to identify, extract, and characterize the various (latent) topics contained by collections of texts. More specifically, topics are identified based on word co-occurrence patterns across a corpus of text documents, where a cluster of words that co-occur frequently across a number of documents constitute a topic. Based on the idea that documents are collections of topics – where a topic represents a probability distribution over words – topic models connect words with similar meanings and differentiate between uses of words with multiple meanings. Each topic is separately meaningful, offering a probability distribution over words which produces a consistent cluster of correlated terms [4, 20–22, 26, 27]. For this study, each comment is considered a distinct document.

When fitting the LDA topic model to a collection of text documents, the analyst needs to specify the number of topics to be identified. This selection generally implies exploration of different solutions to achieving the best fit. We chose eight topics to be detected running the algorithm for 3,000 iterations with the $\sum \alpha = 5$. However, fewer than eight topics are reported because a few topics contain non-English terms and/or have low fitted weights. Finally, topics are not mutually exclusive; member words can be included in more than one topic.

In addition to inferring topics from all comments, we infer a secondary set of topics based on a partition of the corpus according to the most negative and most positive comments, in order to uncover distinct topics associated with opposing ends of the sentiment spectrum. Again, we consider those both containing and not containing the labels; future analyses will consider only those comments containing the selected labels. For negative comments, we consider those with Sentiment (sum) scores ≤ -3 ($n = 4819$) and for positive comments, we consider scores $\geq +2$ ($n = 2517$), due to their being fewer overall positive comments.

3.2.1 Network of Topic Members.

The word membership within topics constitutes a bipartite, topic-word (TW) network. One can 'fold' such a topic-to-word network by multiplying its transpose to itself to obtain a word-to-word network in which linkages represent words co-occurring within the same topic and edge weights (or matrix cells) indicate the extent of the co-occurrence; the matrix calculation is $WW = (TW)^{\mathrm{T}} \times TW$. This transformation can expose those words that span multiple topics; thus, words that co-occur in multiple topics will have weights > 1.

Table 2. Frequency of Labels and Codes (n)

Label	n	*Label*	n
refugee	9020	jihadist/terrorist	1866
refugee crisis	474	criminal	533
migrant	1482	scum	236
migrant crisis	57	muslim	5210
immigrant	1972	islam	2895
immigrant crisis	15	rapefugee	83
		Codes	n
syrian	2456	Refugee	9737
syrian migrant	22	Migrant	1528
syrian refugee	835	Immigrant	1980
asylum seeker	162	Syrian	3073
		Threat	2326

The networks are visualized using the software program Gephi [3] arranged via the Force Atlas algorithm [3] and colored according the Louvain community detection algorithm [5]; nodes having the same color occur in communities in which (roughly) more edges connect members within the community than members of different communities. Nodes and their labels are sized according to their betweenness centrality scores, which captures the extent to which a node lies in the shortest paths between all pairs of nodes.

4. RESULTS

In Table 2, the frequencies (or counts) of each of the labels we investigate are reported. The individual labels (e.g., 'refugee') naturally occur more frequently than through the bigrams (e.g., 'refugee crisis'). As the video addresses both the refugee crisis and the Syrian migration, both 'Refugee' and 'Syrian' umbrella codes are highly represented. However, the predominance of the 'refugee' label and the 'Refugee' code, in comparison to other codes and labels, would suggest a sympathetic tone of the comments in our corpus. As discussed in our introduction, the term refugee denotes individuals fleeing their native countries as a result of armed conflict or persecution. This particular finding raises questions about how this label is used, and what is the valence of the comments employing it. We return to this when discussing our sentiment analysis results.

4.1 Sentiment Analysis Results

The analyses in this subsection focus on the various sentiment scores of the comments containing the selected labels. The first analysis (Figure 1) reveals the unidimensional 'Sentiment' scores and ordering of the labels. The error bars are 95% confidence intervals (CIs), and the colors denote the significance in the difference of means (via t-test) between the colored CIs scores and the mean score of the comments containing the lowest scoring label ('immigrant crisis'). Blue denotes $p < .05$ and green denotes $p < .01$.

Based on this, we note that all average sentiment (and accompanying CIs) reside in the negative region of sentiment, showing that the sentiments in the majority of the comments containing the labels are negative. In fact, the mean sentiment of those comments containing any of the labels ($\mu = -0.97$) are significantly more negative than the mean sentiment from all the other comments

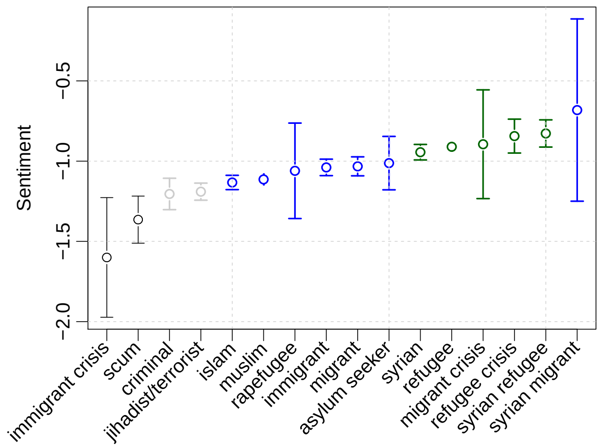


Figure 1. Sentiment of comments containing key concepts.

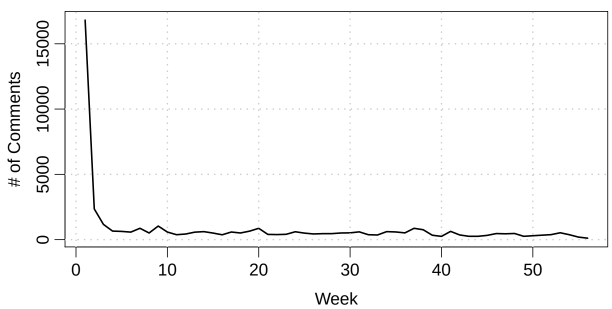


Figure 2. Number of Comments per Week

(i.e., those comments that do not contain any of labels in Table 2), which still bear some nominal level of negativity ($\mu = -0.55$); $t(39922) = 33.3, p < .001$. While naturally, certain labels are expected to elicit higher negativity (such as 'scum', 'criminal', 'jihadist/terrorist', 'rapefugee'), even when controlling for the sentiment in the label itself, others are surprisingly negative, namely, 'immigrant crisis', and 'immigrant'. Bigrams containing 'syrian' appear to elicit the least negative (or most positive) sentiment, suggesting sympathy for their specific situation. Similarly, the labels describing individuals displaced by adverse situations are viewed less negatively than those labels indicating some degree of agency or opportunity (labels containing 'immigrant' or 'migrant').

In Table 3, we compare the difference in sentiment (ΔSentiment) between those comments affiliated with our labels posted in the week of the video's publishing date (September 17, 2015) and those posted in the last two months of the collected data (i.e., September-October 2016). As commenting frequency drops precipitously after the initial week following a video's posting, the second sampling period is larger in order to acquire a sufficient number of samples for testing (see Figure 2). The asterisks accompanying the Δ indicate the significance levels. The s's refer to the mean sentiment during week 0 and weeks ≥ 40; the n's refer to the sample sizes.

Table 3. Comparing early and later sentiments.

	ΔSentiment	$s_{t=0}$	$s_{t\geq40}$	$n_{t=0}$	$n_{t\geq40}$
Labels					
migrant crisis	−0.06	−0.94	−1.00	17	10
syrian	−0.09	−0.90	−0.98	1100	233
immigrant	−0.10	−0.97	−1.07	820	202
muslim	−0.12*	−1.04	−1.16	1745	722
refugee crisis	−0.13	−0.81	−0.95	161	74
refugee	−0.19***	−0.83	−1.02	3310	1111
islam	−0.22**	−1.06	−1.27	911	451
jihadist/terrorist	−0.27***	−1.01	−1.28	509	379
migrant	−0.31**	−0.88	−1.19	619	156
criminal	−0.35*	−1.11	−1.46	204	63
syrian refugee	−0.42*	−0.66	−1.09	325	80
asylum seeker	−0.64*	−0.89	−1.53	75	17
scum	−0.67*	−1.00	−1.67	73	21
Codes					
Immigrant	−0.10	−0.97	−1.07	824	202
Migrant	−0.29**	−0.88	−1.17	634	165
Refugee	−0.20***	−0.81	−1.01	3567	1191
Syrian	−0.17*	−0.84	−1.01	1330	298
Threat	−0.25***	−1.05	−1.30	688	427
comments w/labels	−0.22***	−0.87	−1.08	6425	2298
" w/o labels	−0.15***	−0.47	−0.62	10388	3398

$* = p < .05; ** = p < .01; *** = p < .001$

Firstly, we notice that comments containing all the labels gravitate towards increasingly negative sentiment ($t(4057) = 7.2; p < .001$), as well as all comments not containing any of the labels ($t(6063) = 5.8; p < .001$); that is, all discussion becomes more negative. While this trend may characterize YouTube comment threads in general, the negative trend in sentiment for comments containing labels is more prominent than that of the nominal trend. Also, 'immigrant crisis' does not appear in the table due to low sample sizes in the latter time period, however, we do find that the sentiments surrounding 'immigrant' (and Immigrant) are overall only moderately negative (relative to the other labels) and unchanged through the entire time period. Those labels that change most drastically are typically considered positive or neutral ('asylum seeker' or 'syrian refugee') or strongly negative ('scum', 'criminal'), indicating drastic disaffection for those labels that had already elicited strong sentiment regardless of the valence.

When looking at ordering of change in sentiment for codes (Immigrant, Syrian, Refugee, Threat, and Migrant), the labels associated with the code 'Migrant' exhibits the strongest negative change. This supports the notion that there is less sympathy associated with a characterization of those whose migration are less due to the need to escape from war-torn regions, but rather more likely due to economic hardships or other reasons. Alternative methods such as semantic network analysis may reveal cross-associations between these labels that may appear in the comments.

In Table 4, we regress the sentiment variables of the codes (the broader category indicators), for all comments irrespective of their time stamps. Ordered logit models are employed for the separate sentiment dependent variables (Positive and Negative) and the sum

Sentiment score variable; the Sentiment variable was recoded from -4 to 4 to run from 0 to 8. The notable effects elaborate on some of the earlier findings. The labels containing Immigrant, which as observed earlier, undergoes relatively little change in sentiment; exhibit both most positivity and 2nd most negativity, when the sentiment dimensions are examined separately. Thus, it exhibits high overall emotional Intensity, not only the most negativity (Negative) and relative negativity (Sentiment), but also the most positivity (Positive) and hence overall emotional Intensity. Conversely, comments containing Syrian labels, which exhibited relatively little change as well, contributes to show the least emotional intensity.

When we examine the ordering of Sentiment and Intensity, we find some alignment. For this comparison, we ignore Threat, which is the most negative and provocative code. Both Syrian and Immigrant, respectively, exhibit the lowest and highest coefficients for both Negativity and Intensity, with Asylum Seeker and Migrant exchanging rankings. These orderings can be placed in a larger context in which agency plays a role in determining the sentiment attached to the key labels in each of these broad codes. We revisit this discussion in the Conclusion.

4.2 Topic Modeling Results

Topic modeling can uncover distinct themes within many textual documents, which in our case are individual comments. Each topic comprises words which tend to co-occur across documents. First, we examine the topics of all comments (including those not containing any of the key labels) in Table 5. The first column contains a descriptive label, namely the authors' interpretation of the topic inferred from the words that are members of the topic. The second column shows the Weight measure, indicating relative coherence of the words constituting the topic.

We observe topics directed at regional issues surrounding the crisis (in Germany and Sweden); the conflicts in the Middle East; general commentary on the video itself; extreme, profanity laden antipathy particularly towards Muslims; and religious antipathy directed at Islam.

Table 6 reveals the topics from the top positive and top negative comments. Again, we omit the non-English and relatively very low weighted topics. We note several similarities and differences. Commentary on the video or other comments appear in both positive and negative topics, but positive commentary is distinct in two topics (Video commentary and Exclamatory). The topic of religion also appears in both positive and negative comments, including the descriptive label of 'muslim'. Certain labels such as 'refugee', 'syrian', and 'muslim' feature in both positive and negative comments. Not surprisingly, 'immigrant' features in only the negative comments. In sum, the topic modeling reveals overlap in the usage of some key labels in the topics inferred from the most positive and negative comments of the video.

To explore the overlaps in topics within each set of the most positive and negative comments, we employ network portrayals of topics and their member words, as detailed in the methods section. In Figure 3, we visualize the word-to-word (WW) networks for the most (a) positive and (b) negative comments. For these networks, we include the top 50 frequent words occurring in each topic, yielding a theoretical maximum of $n = 250$ and 300 nodes for, respectively, the positive and negative WW folded networks. The networks statistics of node size (n), edge count ($|E|$), graph density (d), and modularity (q) appear in the subcaptions of each networks.[4]

Through visual observation, as well as the moderate modularity scores (of .411 and .502) of the words that cross topics in both positive and negative networks, 'people' is the most prominent, which is not entirely surprising given the main topic of this paper (the refugee/migrant crisis).

Among the spanning positive comments, we observe positive terms such as 'good', 'accept', 'hope', as well as neutral terms such as 'Europe[an]', 'culture', 'Muslim', 'human', 'fact'. While some of these terms apply to topics relating to the video or other users' comments, they also attest to a positive, humanistic perspective of the refugee crisis.

Some of these spanning words referring to people groups also appear in the set of topic-crossing words for negative comments (such as 'Europe' and 'human') as well as negative words (such as 'afraid', 'hate', 'murder'), indicating that the antipathy-laden terminology is pervading multiple topics of discussion. Furthermore, the extent of cross-topic use is less for the negative topics than for the positive, as indicated by the relatively lower betweenness centrality scores (smaller sizes of the spanning nodes) and the lower modularity q statistic. That is, positivity employs more universal phrasing while the topics for negativity display more distinctiveness. Alternatively, one might argue that commenters find more negative perspectives (and terminology) to the refugee crisis than they do positive ones. Finally, we observe that among the key labels, very few appear prominently among the three topic models – namely, 'refugee', 'immigrant', 'syrian', and 'muslim'; this is likely due to these terms' usage being higher than the other key labels of the situation.

In sum, these results reveal the extent of separation within the more positively-laced and also the more negatively-laced discussions. The comments from the latter harboring the negative sentiments show more focus in its structural positioning of a slightly more generic use of 'people' (based on its betweenness centrality) and the separation of discussion topics. Discussion involving positive comments display greater co-mingling of positive terms as well as more detailed descriptors of people groups.

5. CONCLUSION

Social media platforms, such YouTube, only exist through the continual and growing participation of millions of users, and depend on individual and collective participation and creation of content. Social media responses to societal events, often times characterized by the relative anonymity of personal expression — particularly commenting on YouTube, can lead to empowered and uninhibited public opinion. As such, the use of labels to frame these recent events in Europe can have implications for the lives and safety of refugees, they can undermine public support, steer public opinion,

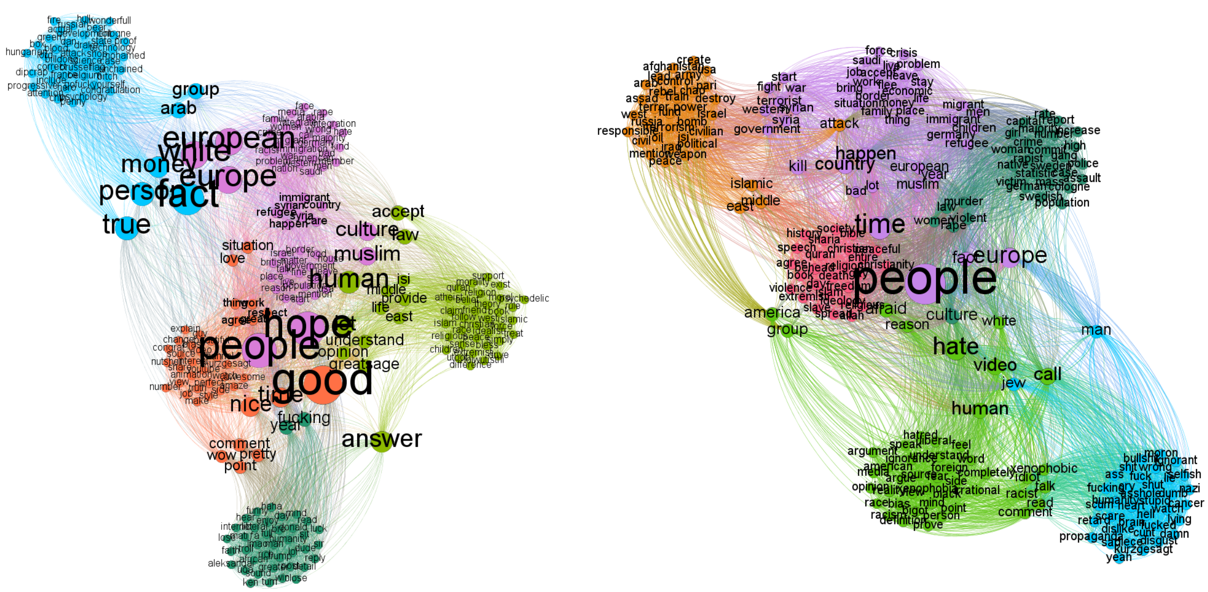
[4]The graph density and modularity statistics range from 0.0 to 1.0. Higher density indicates higher proportion of edge counts over the count of all possible edges and higher modularity indicates distinct communities or clusters.

Table 4. OLS regression of sentiments

	Dependent variable:			
	ordered regression			*OLS*
	Positive	Negative	Sentiment	Intensity
Constant				1.634^{***}
				(0.006)
Asylum Seeker	0.534^{***}	0.567^{***}	-0.127	0.397^{***}
	(0.141)	(0.140)	(0.133)	(0.093)
Refugee	0.478^{***}	0.572^{***}	-0.231^{***}	0.396^{***}
	(0.022)	(0.021)	(0.021)	(0.014)
Syrian	0.239^{***}	0.279^{***}	-0.111^{**}	0.204^{***}
	(0.036)	(0.034)	(0.034)	(0.023)
Migrant	0.347^{***}	0.571^{***}	-0.275^{***}	0.359^{***}
	(0.048)	(0.047)	(0.045)	(0.031)
Immigrant	0.588^{***}	0.801^{***}	-0.329^{***}	0.549^{***}
	(0.042)	(0.041)	(0.040)	(0.027)
Threat	0.400^{***}	0.954^{***}	-0.575^{***}	0.591^{***}
	(0.039)	(0.039)	(0.037)	(0.025)
Observations	46,313	46,313	46,313	46,313
R^2				0.058
Adjusted R^2				0.058
Log Likelihood	-50,008.440	-65,922.160	-77,844.410	
F Statistic				473.357^{***} (df = 6; 46306)

Note #1: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Note #2: Constants for the ordered logit models not reported



(a) Positive topics (n = 222, $|E|$ = 6892, d = .281, q = .502)

(b) Negative topics (n = 235, $|E|$ = 7062, d = .257, q = .411)

Figure 3. Word-word affiliation (for all positive and negative topics).

and influence reactions to this crisis. Frames are never neutral. They define an issue, identify causes, make moral judgments, and shape proposed solutions [42]. The significance of framing lies in the fact that it can affect both individuals and society at large.

Table 5. Topic modeling of all comments

Topic	Wgt	Top Words
Refugees in Germany	0.29	refugee people country europe live war germany problem syrian money
Antipathy	0.26	fuck people fucking shit racist dont comment stupid muslim hate
Video Commentary	0.25	video people comment fact point source argument good opinion bias
Western culture	0.14	people culture european country europe human white western nation society
Middle East conflict	0.13	war country refugee syria isi syrian saudi europe middle east
Refugees in Sweden	0.13	refugee rape muslim crime immigrant sweden europe country population women
Religion	0.11	muslim islam religion people christian kill islamic law quran allah
$\mathcal{L}$ =	-8.28	

At the individual level, exposure to frames may result in altered attitudes, while at the societal level frames can influence processes of political socialization and collective actions [11].

Mostly studied in the context of the mass media, the use of labels as framing instruments has become an integral part of social media and the online world. With this study, we aimed to provide a robust analysis of the dynamics of label use in social media surrounding the recent influx of displaced Middle-Eastern individuals, the emergent patterns of labeling that can cause further disaffection and conflicts, and the sentiments associated with the different labels.

Our analysis of 46,313 comments posted to a single YouTube video on the topic of "The European Refugee Crisis and Syria Explained," showed heavily negative sentiment associated with the selected labels. Furthermore, we also showed that across time, all discussion has become increasingly negative. In particular, labels containing 'migrant' and 'refugee' are distinct in that the latter provoked less negativity, while the former has become associated with more negative sentiment over time. Through topic modeling, we identified the prevailing topics pertaining to more than those raised in the video and include other topics related to the refugee/migrant crisis. A network analysis of the words of each topic revealed extensive overlap in the usage of terms that constitute various discussion of the most positive and negative topics, although negative comments appear slightly more partitioned than the positive ones.

Our study revealed that while there is widespread usage of the various key labels in describing the refugee/migrant crisis and the affected individuals, discussion of the crisis centered around a smaller subset of these labels. Further, the sentiments associated with labels displayed considerable variety in intensity and valence. So, while many of these labels can be argued to be virtually synonymous in a more general context, their framing and interpretation within the context of the crisis can vary considerably – including across time – indicating shifts in online public opinion, but only with specific characterizations of the affected individuals. Prominent labels were integrated into discussion topics found in the overall corpus of the studied comments. How these discussion topics manifest structurally is dependent on the sympathetic (or antipathetic) tone of the discussion. Negatively-laced discussion of the refugee/migrant crisis centered on specific dimensions of racism, concerns or fears of crime, religion, and terrorist activity, while positive discussion highlighted peace and acceptance.

5.1 Discussion of Label Use and Perceived Agency

The results show a distinct pattern when we compare the ordering of the Intensity of sentiment against the changed sentiment of labels over time. Discounting Threat, we see that the Syrian and Immigrant codes exhibit the least amount of changed sentiment, which indicates that their interpretations are relatively more fixed than the other codes. Further, we notice that Syrian exhibits the least Intensity (as well as least negativity overall) and Immigrant exhibits the most Intensity. We offer a hypothetical model that first presumes that most commenters harbor some degree of apprehension or antipathy towards foreigners but then this negative regard is mitigated when the extent of perceived agency is minimal. Here, the concept of perceived agency is employed to suggest the idea that such labels as 'migrant' and 'immigrant' carry an inherent meaning that these individuals have more freedom of choice and suffer from relatively less duress when deciding to leave their countries. As such, Syrian refugees are a specific and perhaps the most sympathetic subgroup of the overall refugees, as envisioned in the minds of commenters, due to their dire situation. Conversely, the Immigrant code (and labels therein) implies that reasons for crossing borders are unknown and perceived to be associated with a much higher degree of agency.

The Migrant also harbors a high degree of agency, but also can be associated with transience or less permanence in residency, which may be viewed with less apprehension in the eyes of commenters. While Refugee and Asylum Seeker imply a lower degree of agency, their motivations for crossing boundaries are less clear than the those associated with the Syrian code. Thus, a model that delineates label interpretation into two dimensions - one of perceived agency and the other of permanence - may explain the rankings we observe in the codes' Sentiment and Intensity. Furthermore, given the conflation of 'refugee' and 'migrant' when paired with 'crisis' in online reports and the news, it is no surprise that both these terms exhibit a relatively large degree of change in sentiment.

This hypothetical model is clearly imperfect at the moment, partly due to the large changes in sentiment for the specific labels of 'syrian refugee' and 'asylum seeker'. Further investigation is warranted for uncovering how the interpretation of these labels may be positioned through the latent dimensions of perceived agency and permanence of outsiders as well as fluidity of labels, perhaps due to the manner in which they are employed by (social) media.

Table 6. Topic modeling of positive and negative comments

Topics from most positive comments (5 relevant listed, $\mathcal{L} = -8.03$)		
Topic	Wgt	Top Words
Video commentary	0.45	video great good love nice people guy channel job work
Refugee	0.21	people country live good europe refugee life syrian year place
Exclamatory	0.14	good comment nice lmao read wow pretty lol hope point
Acceptance	0.14	people refugee country europe love muslim syrian accept nation american
Religious peace	0.08	religion muslim islam christian human peace law culture sense friend
Topics from most negative comments (6 relevant listed, $\mathcal{L} = -8.05$)		
Topic	Wgt	Top Words
Middle East conflict	0.51	refugee people country europe live syrian problem war syria european
Antipathy	0.38	fucking fuck comment video shit idiot hate people bullshit stupid
Criminal	0.26	rape crime muslim sweden refugee women europe commit immigrant year
Religious conflict	0.22	muslim religion islam people kill christian culture hate law islamic
Racism	0.18	people fear fact hate point source racist argument white group
Terrorism/ISIS	0.14	terrorist war attack isis syria kill middle east start group

5.2 Limitations and Future Research

The dataset analyzed in this study pertains to a single, English language YouTube video, and thus it limits the extent to which we can generalize our findings in several ways. Firstly, the single video was sympathetic in tone to the European refugee crisis, which may introduce a certain bias to the type of user responses it triggered or to the audience it attracts. As such, it can be argued that the results presented in this study may not be representative of the overall online discussions of the European refugee crisis. Naturally, a broader future study will include discussions surrounding additional videos. Secondly, while the narration of the video is in English, many of the countries affected by the migrant crisis are not anglophone. So, the framing of the crisis in these countries may differ. Therefore, a more comprehensive understanding of the overall perception of the refugee crisis in Europe would entail the inclusion of videos that employ the languages of those countries.

As it is the case with most online data, a moderate portion of language found in online textual comments are written in informal English or grammatically incorrect. Our data thus contained a blend of abbreviations, slang, and context specific terms. While various data pre-processing methods are available, large corpora of grammatically incorrect text remains a challenge in the study of user generated web data. User generated data, such as the YouTube comments we investigated, also contains a large number of malformed words and colloquial expressions (e.g, 'looov', 'luv', 'gr8', 'lol' etc.). Such informal (English) text poses certain challenges for sentiment analysis methods. More specifically, sentiment analysis often employs lexicon-based approaches, which use lexicons of words weighted with their sentiment orientations to determine the overall sentiment of a given text. Although these approaches have been proven highly effective for conventional text [38], they tend to be ill suited for online textual data.

The sentiment analysis method employed in this article comes from Thelwall et al. [56] and uses a human-coded lexicon of words and phrases specifically built to work with online social data. The proposed algorithm, SentiStrength, utilizes this human-coded lexicon to identify the sentiment strength of informal text (e.g., tweets, status updates, YouTube comments). Although SentiStrength has proven relatively accurate and consistent in analyzing social media data, its results remain confined to the fixed set of words that appear in its lexicon. This may pose problems when dealing with online textual data, where new expressions and jargon constantly emerge.

The significant results we presented prompt us to further explore our data through topic modeling of only those comments containing the selected labels to reveal the most prominent topics associated with these labels. Our future plans include a more in-depth analysis of the salient words surrounding the labels via semantic network analysis; a more in-depth temporal or time series analysis to expose the dynamics of label use; and alternative sentiment scores based on individual words rather than the most extreme sentiment-laden terms in a comment. Lastly, we plan on extending this work by including an even more extensive corpus of YouTube comments into our analysis, such as other videos pertaining to the refugee/migrant crisis.

As social media becomes more prevalent, incurring higher levels of participation and creation of user-generated content, studies of online opinions and discussions, such as this paper, become increasingly valuable by offering insights into the nature and direction of focal discussion themes and public sentiment surrounding those themes. Still, the nature of these discussions and expressions of opinion may be strongly dependent on the characteristics of the platforms in which they occur, such as the level of anonymity afforded in the interactions among the users.

REFERENCES

[1] M. Augoustinos and C. J. Quinn. 2003. Social categorization and attitudinal evaluations: Illegal immigrants, refugees, or asylum seekers? *New Review of Social Psychology* 2 (2003), 29–37.

[2] Gabriela Baczynska and Robert-Jan Bartunek. 2016. Europe's free travel will end unless Turkey halts migrant flow, officials say. Online. (25 February 2016). Retrieved May, 30, 2017 from http://www.reuters.com/article/us-europe-migrants-idUSKCN0VY2M8

[3] M. Bastian, S. Heymann, and M. Jacomy. 2009. Gephi: an open source software for exploring and manipulating networks. *Proceedings of the Third International ICWSM Conference*, 361–362.

[4] D. M. Blei, A. Y. Ng, and M. I. Jordan. 2003. Latent dirichlet allocation. *Journal of Machine Learning Research* 3 (2003), 993–1022.

[5] V. D. Blondel, J-L Guillaume, R. Lambiotte, and E. Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008, 10 (2008), 1–12.

[6] T. Brader, N. A. Valentino, and E. Suhay. 2008. What triggers public opposition to immigration? Anxiety, group cues, and immigration threat. *American Journal of Political Science* 52, 4 (2008), 959–978.

[7] P. R. Brewer. 2001. Value words and lizard brains: Do citizens deliberate about appeals to their core values? *Political Psychology* 22, 1 (2001), 45–64.

[8] C. Canali, M. Colajanni, and R. Lancellotti. 2010. Characteristics and evolution of content popularity and user relations in social networks. In *Computers and Communications (ISCC), 2010 IEEE Symposium on*. IEEE, 750–756.

[9] G. Chatzopoulou, C. Sheng, and M. Faloutsos. 2010. A first step towards understanding popularity in youtube. In *INFOCOM IEEE Conference on Computer Communications Workshops, 2010*. IEEE, 1–6.

[10] D. Chong and J. N. Druckman. 2007. Framing theory. *Annual Review of Political Science* 10 (2007), 103–126.

[11] C. H. de Vreese. 2005. News framing: Theory and typology. *Information Design Journal + Document Design* 13, 1 (2005), 51–62.

[12] J. Diesner, A. Aleyasen, C. Chin, S. Mishra, K. Soltani, and L. Tao. 2015. ConText: Network Construction from Texts. (2015). Available at: http://context.lis.illinois.edu.

[13] J. N. Druckman. 2011. What‘s it all about? Framing in political science. In *Perspectives on framing*, G. Keren (Ed.). Psychology Press, New York, 279–302.

[14] Ben Endley. 2014. Migrants ARE swamping parts of Britain, says David Blunkett after Michael Fallon comment. Online. (28 October 2014). Retrieved May, 30, 2017 from http://www.express.co.uk/news/uk/528428/David-Blunkett-migrants-swamping-Britain

[15] R. M. Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of Communication* 43, 4 (1993), 50–58. doi:10.1111/j.1460–2466.1993.tb01304.x.

[16] Willa Frej. 2015. "You're Not Welcome Here," French Mayor Tells Syrian Refugees. Online. (14 September 2015). Retrieved May, 30, 2017 from http://www.huffingtonpost.com/entry/french-syrian-refugees_us_55f6f734e4b077ca094fbefb

[17] W. A. Gamson and G. Wolfsfeld. 1993. Movements and media as interacting systems. *The Annals of the American Academy of Political and Social Science* (1993), 114–125.

[18] P. Gill, M. Arlitt, Z. Li, and A. Mahanti. 2007. Youtube traffic characterization: a view from the edge. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*. ACM, 15–28.

[19] E. Goffman. 1974. *Frame analysis: An essay on the organization of experience*. Harvard University Press, New York.

[20] T. Griffiths and M. Steyvers. 2002. A probabilistic approach to semantic representation. In *Proceedings of the 24th Annual Conference of the Cognitive Science Society*. 381–386.

[21] T. Griffiths and M. Steyvers. 2003. Prediction and semantic association. In *Neural Information Processing Systems 15*, S. Becker, S. Thrun, and K. Obermayer (Eds.). MIT Press, Cambridge, MA, 11–18.

[22] T. Griffiths and M. Steyvers. 2004. Finding scientific topics. In *Proceedings of the National Academy of Sciences*. National Academy of Sciences, 5228–5235.

[23] K. Hallahan. 1999. Seven Models of Framing: Implications for Public Relations. *Journal of Public Relations Research* 11, 3 (1999), 205–242. https://doi.org/doi:10.1207/s1532754xjprr1103_02

[24] A. Haynes, E. Devereux, and M. Breen. 2008. Public exercises in othering: Irish print media coverage of asylum seekers and refugees. In *Facing the Other: Interdisciplinary Studies on Race, Gender and Social Justice in Ireland.*, B. Farago and Sullivan. M. (Eds.). Cambridge Scholars Press, Newcastle, 162 – 181.

[25] M. Heckner and C. Wolff. 2009. Towards Social Information Seeking and Interaction on the Web. In *Information: Droge, Ware oder Commons? Wertschöpfungs- und Transformationsprozesse auf den Informationsmärkten.*, R. Kuhlen (Ed.). Vol. 50. vwh Verlag Werner Hülsbusch, Boizenburg, 235–241.

[26] T. Hofmann. 1999. Probabilistic latent semantic indexing. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 50–57.

[27] T. Hofmann. 2001. Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning* 42, 1 (2001), 177–196.

[28] K. Horsti. 2007. Asylum seekers in the news: Frames of illegality and control. *Observatorio Journal* 1 (2007), 145–161.

[29] S. Iyengar. 1987. Television News and Citizens' Explanations of National Affairs. *American Political Science Review* 81 (1987), 815–831. doi:10.2307/1962678. Issue 3.

[30] G. M. Jones and B. B. Schieffelin. 2009. Talking text and talking back: "My BFF Jill" from boob tube to YouTube. *Journal of Computer-Mediated Communication* 14, 4 (2009), 1050–1079.

[31] D. Kahneman and A. Tversky. 1984. Choices, values, and frames. *American Psychologist* 39, 4 (1984), 341–350.

[32] R. Kirk and D. Schill. 2011. A digital agora: Citizen participation in the 2008 presidential debates. *American Behavioral Scientist* 55, 3 (2011), 325–347.

[33] B. R. Knoll, D. P. Redlawsk, and H. Sanborn. 2011. Framing Labels and Immigration Policy Attitudes in the Iowa Caucuses: "Trying to Out-Tancredo Tancredo". *Political Behavior* 33, 3 (2011), 433–454.

[34] K. Kousha, M. Thelwall, and M. Abdoli. 2012. The role of online videos in research communication: A content analysis of YouTube videos cited in academic publications. *Journal of the American Society for Information Science and Technology* 63, 9 (2012), 1710–1727.

[35] P. G. Lange. 2007. Commenting on comments: Investigating responses to antagonism on YouTube. In *Society for Applied anthropology conference*, Vol. 31. 2007.

[36] P. G. Lange. 2007. Publicly private and privately public: Social networking on YouTube. *Journal of Computer-Mediated Communication* 13, 1 (2007), 361–380.

[37] S. E. Lillie. 2008. Diffusion of innovation in the age of YouTube. *American Journal of Preventive Medicine* 34, 3 (2008), 267.

[38] B. Liu. 2010. Sentiment analysis and subjectivity. In *Handbook of Natural Language Processing*, N. Indurkhya and F. J. Damerau (Eds.). Chapman and Hall/CRC, London, 627–666.

[39] A. Madden, I. Ruthven, and D. McMenemy. 2013. A classification scheme for content analyses of YouTube video comments. *Journal of Documentation* 69, 5 (2013), 693–714. https://doi.org/10.1108/JD-06-2012-0078

[40] A. Kachites McCallum. 2002. MALLET: A Machine Learning for Language Toolkit. (2002). Available at: http://mallet.cs.umass.edu.

[41] M. C. Milliken and S. O'Donnell. 2008. User-generated online video: The next public sphere?. In *Technology and Society, 2008. ISTAS 2008. IEEE International Symposium on*. IEEE, 1–3.

[42] S. O'Neill, H. T. P. Williams, T. Kurz, B. Wiersma, and M. Boykoff. 2015. Dominant frames in legacy and social media coverage of the IPCC Fifth Assessment Report. *Nature Climate Change* 5, 4 (2015), 380–385.

[43] Z. Pan and G. M. Kosicki. 1993. Framing analysis: An approach to news discourse. *Political Communication* 10, 1 (1993), 55–75. https://doi.org/10.1080/10584609.1993.9962963

[44] J. C. Paolillo. 2008. Structure and network in the YouTube core. In *Proceedings of the 41st Annual HICSS conference*. IEEE, 156–156.

[45] A. J. Porter and I. Hellsten. 2014. Investigating participatory dynamics through social media using a multideterminant "frame" approach: the case of Climategate on YouTube. *Journal of Computer-Mediated Communication* 19, 4 (2014), 1024–1041.

[46] Rieder, B. 2015. YouTube Data Tools. Available at: https://tools.digitalmethods.net/netvizz/youtube/. (2015).

[47] C. Roggeband and M. Verloo. 2007. Dutch women are liberated, migrant women are a problem: The evolution of policy frames on gender and migration in the Netherlands, 1995 - 2005. *Social Policy & Administration* 41, 3 (2007), 271 – 288.

[48] C. Roggeband and R. Vliegenthart. 2007. Divergent framing: The public debate on migration in the Dutch parliament and media, 1995–2004. *West European Politics* 30, 3 (2007), 524–548.

[49] M. J. Rohan. 2000. A rose by any name? The values construct. *Personality and Social Psychology Review* 4, 3 (2000), 255–277.

[50] H. Rojas, D. V. Shah, J. Cho, M. Schmierbach, H. Keum, and H. Gil-De-Zuñiga. 2005. Media dialogue: Perceiving and addressing community problems. *Mass Communication & Society* 8, 2 (2005), 93–110.

[51] P. Schultes, V. Dorner, and F. Lehner. 2013. Leave a Comment! An In-Depth Analysis of User Comments on YouTube. *Wirtschaftsinformatik* 42 (2013).

[52] David Shariatmadari. 2015. Swarms, floods and marauders: the toxic metaphors of the migration debate. Online. (10 August 2015). Retrieved May, 30, 2017 from https://www.theguardian.com/commentisfree/2015/aug/10/migration-debate-metaphors-swarms-floods-marauders-migrants

[53] S. Siersdorfer, S. Chelaru, W. Nejdl, and J. San Pedro. 2010. How useful are your comments? analyzing and predicting youtube comments and comment ratings. In *Proceedings of the 19th international conference on World wide web*. ACM, 891–900.

[54] P. Sobkowicz, M. Kaschesky, and G. Bouchard. 2012. Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web. *Government Information Quarterly* 29, 4 (2012), 470–479.

[55] M. Thelwall. 2013. Heart and soul: Sentiment strength detection in the social web with SentiStrength. *Proceedings of the CyberEmotions* (2013), 1–14.

[56] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas. 2010. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology* 61, 12 (2010), 2544–2558.

[57] A. Tversky and D. Kahneman. 1985. The framing of decisions and the psychology of choice. In *Environmental Impact Assessment, Technology Assessment, and Risk Analysis*, V. T. Covello, J. L. Mumpower, P. J. M. Stallen, and V. R. R. Uppuluri (Eds.). Springer, Berlin, 107–129.

[58] Wikipedia. 2015. European migrant crisis. Online. (2015). Retrieved May, 30, 2017 from https://en.wikipedia.org/wiki/European_migrant_crisis