RC25672 (IRE1711-013) November 14, 2017
Computer Science

IBM Research Report

Trends and Topics in Decision Tweets

Lea A. Deleris
IBM Research
IBM Dublin Technology Campus, Bldg. 3
Damastown Industrial Estate
Mulhuddart
Dublin 15, Ireland

Research Division
Almaden – Austin – Beijing – Cambridge – Dublin - Haifa – India – Melbourne - T.J. Watson – Tokyo - Zurich

Trends and Topics in Decision Tweets

Léa A. Deleris

Abstract:

The objective of this research is to explore to what extent Twitter data is pertinent to decision analysis research. Do tweets contain information that is directly usable for decision making research? Could they lead decision researchers to develop wide-scale insights into people's understanding, behavior and opinions related to decision making? In this initial investigation, we present a descriptive analysis of more than 350,000 tweets associated with keywords "decision" and "making". We first report on the statistical traits of their authors along with the associated time patterns. We then examine the content of the tweets through the use of topic models. By providing an initial overview of the trends and topics of tweets focused on decision making, our intent is to enable researchers in decision analysis to better understand this data source and use the findings presented here to form an opinion as to whether they should consider it for their studies.

1. Introduction

Social media with its constant spontaneous outpour of information from millions of users appears like a mane for any research that focuses on understanding human behavior (Golder and Macy 2012). Naturally, information is not as consumable as data gathered in a regulated experimental setting but it has the benefit of being more obviously representative of people's behavior, values and concerns "in the wild". We are particularly interested in this research in data generated from microblogs, places where users can post and share short statements, e.g., Facebook, LinkedIn and Twitter. While microblogs are outlets for comments on typical news media topics, they also offer insights into users' personal lives and opinions (Zhao et al., 2011).

There has been multiple studies focusing specifically on the use of topic models on Twitter data. In particular, there has been research into optimal models and approaches to apply topic models to tweets (Hong and Davison 2010; Vosecki et al. 2013). Ramage, Dumais and Liebling (2010) sought to provide a richer representation of content of tweets so as to increase performance of recommendation tasks (i.e., posts to read or users to follow).

While we make use of topic models in this research, our objective is different from the above studies which aimed at providing universal tools for the analysis of tweets. We propose in this paper to analyze the content of a specific subset of tweets, those associated with decision making. We are interested in particular in both the trends of such tweets compared to other tweets in terms of timing, frequency, retweeting behavior, and also in terms of characteristics of their authors. We are also exploring which topic of discussions are associated with such tweets. Perhaps will our findings resonate with some existing research directions in the decision science community? Perhaps will they also generate novel research questions or experiments in decision science? This introductory paper does not offer definite answers to these questions, rather seeks to provide deeper information about decision tweets so as to allow researchers to make an informed decision about investigating Twitter data.

2. Semantics of Decision Making

One essential step of our process is the selection of tweets that pertain to decision making. Ideally, one would need an expert in decision making to go through all possible tweets and indicate which subset is relevant and which subset is not. In addition, prior to that, such expert would also have to clarify what is meant by decision making.

In practice, one needs to resort to simpler ways to select relevant tweets, starting with using keywords. In this approach the difficulty is finding a balance between precision (proportion of tweets retrieved that are relevant) and recall (proportion of relevant tweets that is retrieved through the process). Note also that those metrics could only be assessed if we already know what is relevant, making them impractical here given the volume of tweets, but they are useful concepts to think about the performance of information retrieval and extraction tasks.

In this specific research effort, we are interested in decision making as the process of making a decision, in that sense being fully aligned with the definition provided by the WordNet thesaurus[1] (Princeton University, 2010) for decision making : "the cognitive process of reaching a decision". The challenge lies in defining which set of keywords would best yield our target tweets. As we dive into the discourse of the population at large, we need to allow for a different usage of the same words than those of the decision analysis community. Consider for instance the gap between our understanding of the word preferences and how it used casually. In this paper, we have sought a simple approach by relying solely on the terms "decision" and "making" though we allow for those words not to be side by side nor in the same order. As shown through the definition above, the general understanding of these terms fits with our intent.

[1] http://wordnetweb.princeton.edu/perl/webwn

Granted, there will be tweets that deal with decision making though they do not contain any of those words, thus affecting our recall performance. An alternative approach could seek to expand on the keywords through the use of synonyms. We briefly describe here options for such a keyword expansion. One natural direction is to make use of thesauri, WordNet being broadly used in the Natural Language Processing research and development community. The difficulty with such resources comes from the polysemy of words. For instance, the noun form "deciding" is a synonym for "decision making" though its verb form has many additional senses (such as bring to an end : "The judge decided the case in favor of the plaintiff"). As we cannot constraint on part of speech (e.g., verb, adjective, noun) when querying via keywords, adding "deciding" to the list of keywords may increase noise more than it helps in retrieving additional relevant tweets.

Beyond relying on thesauri, another possible direction for keyword expansion is semantic similarity services that provide a list of nearest neighbors to a specific word based on a statistical analysis of its usage (i.e., words are similar when they are frequently used in the same context defined as neighboring words) within a given corpus. One well-known such approach is based on the word2vec algorithm (Mikolov et al. , 2013). As these models can be trained on a corpus, they are able to capture similarity of usage within a certain context. However, training requires vast amount of input data, though few pre-trained models available. Table 1 provides the list of neighbors to the word "decision" relying on a model trained on the Google News corpus. With such models, care needs to be taken to set an appropriate threshold for inclusion. Finally, those systems typically do not handle phrases, i.e., multiword expressions such as "decision making" well.

Table 1 : Nearest neighbors to the word "decision" using a word2vec model trained on the Google News Corpus

Word	Cosine distance
decisions	0.527
make	0.471
makes	0.451
made	0.407
unnecessary	0.392
deliberation	0.386
judgments	0.385
arbitral	0.383
choosing	0.375
unanimous	0.373

3. Data Gathering and Processing

To assemble our dataset of tweets focused on decision making, we made use of the Twitter Decahose, a 10% random sample of all published tweets, which we accessed using IBM Bluemix service Insights for Twitter[2]. As previously discussed, we selected all tweets that corresponded to the following query: "Decision AND Making AND lang:en" over a time period of about 2 years. This corresponds to all tweets that contain both terms "decision" and "making" and that are tagged by Twitter as being written in English.

For the period April 2014 (start date of the Decahose at the time of the data access) to April 2016 (date of the data access), we were able to analyze 368259 tweets corresponding to this query. Here are a few examples of tweets that were retrieved: *"If you get salad, excuse yourself from smart decision making"*, *"Senior year is the year of decision making, and I'm terrible at it"*, *"Emotions have a crazy way of screwing with your decision making"*, *"Thinking too much about*

[2] https://console.ng.bluemix.net/docs/services/Twitter/Twitter_overview.html#about_Twitter

a decision can lead to not making a decision at all", "The biggest decision I'm making today is whether I'm going to watch Man utd vs Bayern or Athletico vs Barcelona", "@<TWITTER_USER> *BTW thank you for making the brave decision to end your music career".*

Besides regular tweet information, each tweet was additionally tagged with information about the sentiment expressed in the body of the tweet in addition to generic information such as date, and information about author location, gender, marital and parental status. Figure 1 provides an example of such a tweet formatted in JSON for our analysis.


```
{
  "_id": "tag:search.twitter.com,2005:450785291356610560",
  "_rev": "1-8fea221db2eddf6b0d049ce56a01d288",
  "tweet_id": "tag:search.twitter.com,2005:450785291356610560",
  "polarity": "POSITIVE",
  "author": {
    "gender": "unknown",
    "isParent": "unknown",
    "id": "BlackSpyda73",
    "isMarried": "unknown"
  },
  "location": {
    "country": "",
    "city": "",
    "state": ""
  },
  "body": "I am constantly making the decision between being a good person or being a sarcastic one",
  "postedTime": "2014-04-01T00:02:43.000Z",
  "link": "http://twitter.com/BlackSpyda73/statuses/450785291356610560"
}
```


Figure 1: Decision Making Tweet JSON Example

We undertook basic post-processing of the tweets to expand information related to timestamp by extracting associated year, month, day, day of the week. We also added a feature to identify retweets (those whose body started with "RT @<username>" at the beginning).

In order to compare decision tweets to regular tweets, we assembled in the same manner a comparison set of tweets of the same size. Specifically, for each possible date appearing in our initial tweet dataset, we have randomly selected the same number of tweets from the decahose, without constraining besides being written in English, leading to a second set of 368259 comparison tweets. We refer to our main dataset of tweets as the “decision tweet” dataset and to the comparison one as the “ordinary tweets” dataset.

4. Statistical Analysis

Missing values prevent us from making use of parental and marital status. For the field isParent, only 1989 answers were non-empty, all being positive. Similarly for the field isMarried, we counted 637 non-empty answers, all of them also positive. We did not make use of the location information which is seldom available and as a free text entry, often inconsistent.

Overall, the useful descriptive fields for statistical characterization are:

- Polarity which can take values neutral, positive, negative and ambivalent
- Gender (male and female)
- Day, Month and Year derived from the timestamp
- DayOfWeek derived from the timestamp
- IsRetweet (binary) derived from the tweet body

None of the fields have missing values. Gender can be unspecified (“unknown”) as Twitter does not impose on users to provide this information when they register. Polarity can be “EMPTY” if the sentiment analysis algorithm could not determine any sentiment, though it concerns only 5000 tweets in our decision tweet dataset. The polarity label is automatically generated from the tweet extraction service, providing no opportunity for setting up a threshold of acceptance for the non-neutral labels. To help readers gauge what polarity represents, Table 2 provides examples of tweets associated with each of the four labels.

Table 2 : Example of Tweets associated with Polarity Labels

Polarity	Example of tweets
POSITIVE	Congrats to <TWITTER_USER> on finally making her decision to go to Missouri State!! I don't know if I'm making the best decision; but it feels like a step in the right direction.
NEUTRAL	I hope I'm making the right decision... Making the decision to go is the hardest part.
NEGATIVE	Always afraid of making the wrong decision. Making Maths our last exam was possibly the worst decision ever.
AMBIVALENT	Feel really bad for making that decision that make the one I love down.

Volume and Timing

We start by looking at the level of activity related to decision tweets overtime, analyzing the frequency and volume of tweets along with associated retweet rate. Figure 2 reports the daily number for tweets for the covered period (left axis) along with the associated percentage of retweets. Overall, the number of daily tweets oscillates around 450 with a median value of 442, for a minimum over the period of 208 daily tweets and a maximum of 4474. In terms of volume of retweet, 40% of the decision tweets are retweets (though they can be retweets of different tweets) with a median of 34.5% , a minimum of 2% and a maximum of 82%.

We investigated the main outliers in Figure 2 as summarised in Table 3. Most of the large volume days are the consequence of tweet spamming where some tweets are published many times without being explicitly presented as Retweet. An example of such spam tweet is "*The lingering uncertainties make decision-making extraordinari... More for Pisces* http://t.co/xrzTv4WjcD". Such noisy tweets mention astrology signs but need not though they typically represent unsolicited advice for instance "*There's nothing wrong with your decision-making skills today*".

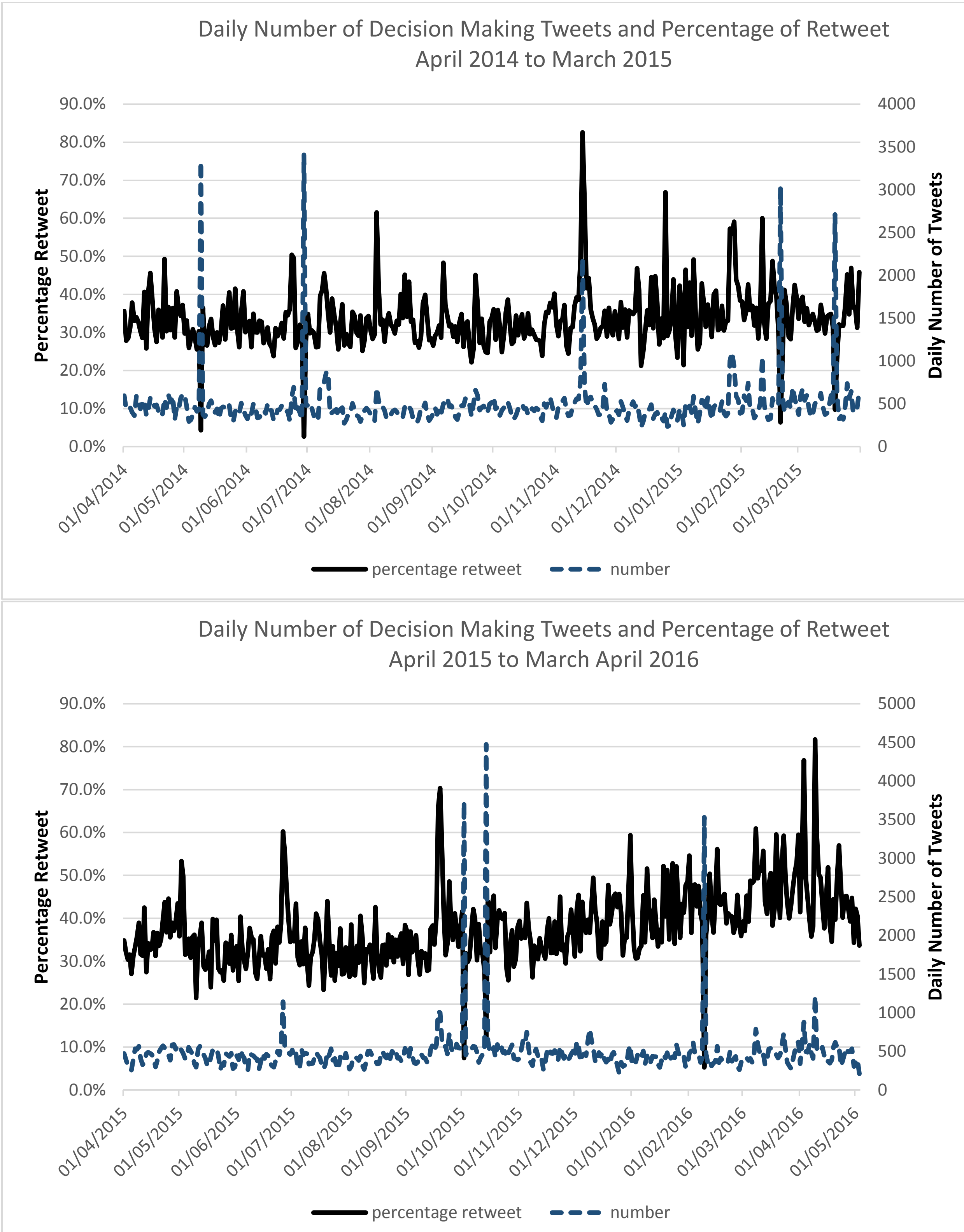

Figure 2: Daily Volume of Decision Tweet and Retweet Rate
Note: The graph has been split solely to increase readability

There is one exception in our analysis corresponding to November 14th 2014 where the volume of decision tweets was driven by the overwhelming retweet of a football (soccer) related comment "*I knew there was problems with decision making at FIFA when I didn't make the ballon d'Or shortlist*". We removed the identified noisy tweets from the daily counts and observe that the adjusted number of tweets for each outlier day falls back into the expected range (last column of Table 3). Overall, despite our expectation that outlier days would be indicators of a specific *societal* interest in decision making, we found no evidence of this behavior in our current study, except possibly of the importance of sports among people's interests.

Table 3 : Analysis of Outlier Days

Date	Number of Tweets for that Day	Proportion of retweet	Diagnostic	Adjusted number of tweets
09/05/2014	3328	4.3%	astrology spam	465
29/06/2014	3415	2.6%	spam not astrology	299
14/11/2014	2198	82.6%	specific tweet	532
20/02/2015	3031	6.4%	astrology spam	319
19/03/2015	2710	9.7%	astrology spam	598
02/10/2015	3734	7.5%	astrology spam	626
14/10/2015	4474	12.3%	astrology spam	930
09/02/2016	3530	5.3%	spam not astrology	510

We also looked at the proportion of decision tweets relative to the total volume of English language tweets being published each day. Decision related tweets represent a small proportion of the global volume, on average 0.004% of the global tweets are decision-making related, varying between 0.001% and 0.04% (corresponding to outlier day October 14th 2015 due to an astrology spam). Timewise, we observe a relative stability of the volume of decision tweets over time as displayed on Figure 3. As the total number of tweets is fairly stable (around

10M to 12M tweets daily) over the period that we consider, we thus conclude that we are in some “steady-state” in terms of decision tweet publishing behavior.

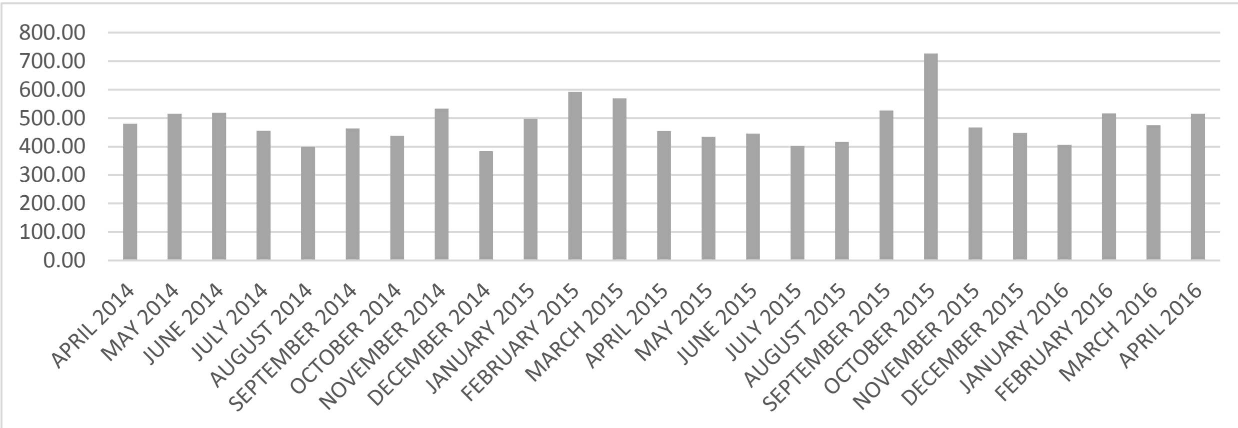

Figure 3: Daily Volume of Decision Tweet Averaged by Month

Regarding day of the week, decision tweets are not evenly distributed, they are more likely to occur Tuesday to Friday (around 16% for each of those days), 14% chance to fall on a Monday, 11% on a Saturday and 12% on a Sunday.

Gender and Polarity

We now analyze the characteristics of the decision tweets and their authors through a specific focus on Gender and Polarity (See Figure 4).

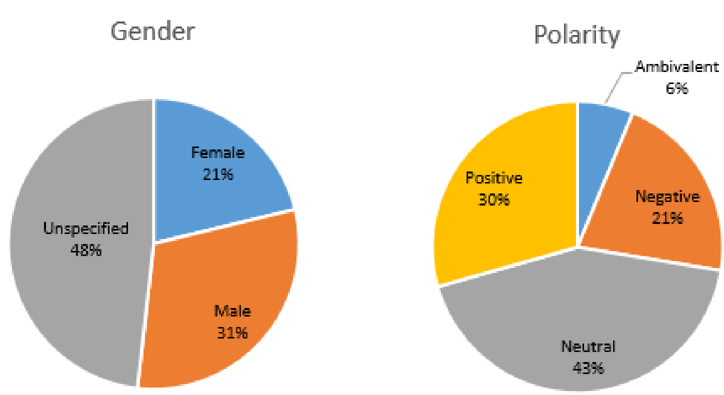

Figure 4: Summary of Decision Tweet Characteristics

Gender wise, about half of the tweets are unspecified, however, for the other half, we count a larger proportion of male with a ratio of about 1.5 of male to female authors. Regarding polarity, normalized for "EMPTY" entries, 43% of decision tweets are tagged as neutral, 30% as positive, 21% as negative and only 6% are tagged as ambivalent. In themselves, those metrics are moderately insightful though the gender distribution does suggest that a difference between male and female in terms of focusing tweets on decision making.

To further understand decision tweets, we performed cross-tabulation for most of our descriptive fields along with Pearson's chi-squared tests of independence which were performed using Matlab. Specifically, we performed the set of analyses on two input sets, the first one corresponding to the complete decision tweet dataset, the second one where all retweets have been removed. Overall, the results are similar whether we apply to either dataset. We provide the cross tabulation data with associated p-values in the Appendix.

For polarity, those tests reveal that authors that specify their gender tend to be less neutral than the ones that do not, men are more positive and women slightly more negative when they tweet about decision making. Also, over time (by year), decision tweets have become more neutral (from 40% in 2014 to 46.5% in 2016) leading to a lower proportion of tweets being negative or positive in 2016 than in 2014. Figure 5 provides the distribution of polarity tag for each day of the week, revealing a greater propensity to express neutral opinions about decision during week days, a spike in negativity on Fridays and more positivity on Sundays.

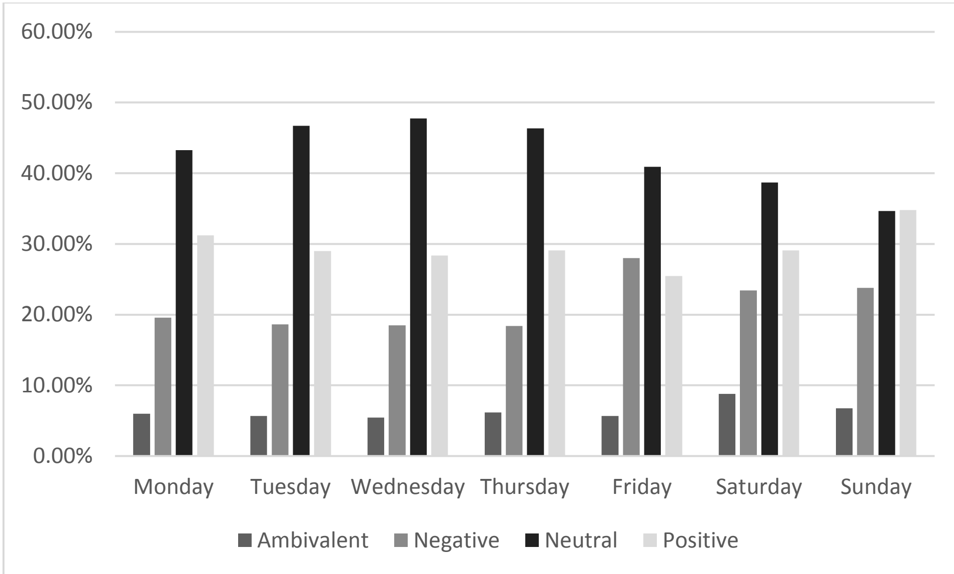

Figure 5: Distribution of Polarity by Day of the Week.

Comparison with Ordinary Tweets

Our last set of statistical analyses pertains to the comparison with the ordinary tweets dataset, so as to determine whether tweets and authors of tweets focusing on decision have some specificities from the general population of tweet authors. We analyze differences in terms of polarity, gender and retweeting behavior. We chose to look at the following comparisons configurations:

- All decision tweets vs. all ordinary tweets
- All decision tweets but with no retweet vs. all ordinary tweets without retweets

As in the previous section, we rely on chi-squared tests to assess the significance of the observed differences. Results show that for all fields (polarity, gender and retweeting behavior), there is a significant difference between decision tweets and ordinary tweets, regardless of the comparison configuration. All results are reported in the Appendix.

Overall, decision tweets are more opinionated than ordinary tweets. Around 42% of decision tweets are neutral versus about 55% for ordinary tweets. The difference comes from ambivalent tweets (6% in decision tweets vs. 3% in ordinary tweets) and negative sentiment (above 20% for decision tweets vs. 13% for ordinary tweets). Decision tweets are also ever so slightly more positive. As we have seen in previous analyses, decision tweets are more strongly associated with male, 60% are from men versus 50% in ordinary tweets. Finally, retweet behavior is also different, while around 27% of the ordinary tweets are retweets, this number increases to 35% for decision tweets. The fact that users are more interested in re-sharing tweets about decisions than a random ordinary tweet indicates that decision-making is a relatively important topic of interest for this population.

5. Topic Models

Overview

We now turn to the exploration of the content of the tweets. The natural first step consists in discovering the themes of the tweets though the application of topic models. We apply one of the standard approach for this task, namely Latent Dirichlet Allocation (Blei, Ng and Jordan 2003). Topic model is an unsupervised approach to discover, from a set of input documents, what are the topics discussed and which topics are discussed in which documents. More specifically, considering the input vocabulary (array of all words found in the input documents) and a target number of topics *T* of topics to be uncovered, each topic is associated with a multinomial distribution over the vocabulary and each document is associated with a multinomial distribution on the identified *T* topics.

We organized our dataset of tweets in documents as follow: (i) each individual tweet corresponds to a document or (ii) all tweets are aggregated together to form a single document, which corresponds to the recommendation from the literature (Hong and Davison 2010). As in the previous section on statistical analysis, we investigate the effect of considering either full set of (cleaned) tweets or only the subset without the retweets, thereby leading to four experimental configurations to create the input to the topic model.

As we focus on the content of the tweets, it is necessary to perform some preprocessing on the body of the tweets prior to learning topic models. Specifically, we remove all retweet handles (RT), we replaced Twitter username with "Twitter_user", internet links with "link" and ampersand sign (&) with "and". We make use of the normalization dataset from Li and Liu (2015) to correct some of the misspellings in the body such as "2nite" or "4got". In addition, based on the findings from the statistical section, we remove all tweets containing references to the zodiac signs from western astrology. Finally, we remove common words of the English language (e.g., the, who, a, I, do) through the use of a stop word list to which we manually add the terms "decision" and "making" for the tweets in the decision dataset.

To implement LDA, we rely on the MALLET library (McCallum 2002), with standard values for hyper parameters ($\alpha=1/T$ for the symmetric Dirichlet distribution of documents over topic and $\beta=0.01$ for the symmetric Dirichlet distribution of topics over words and 2000 iterations). We vary the topic size parameter (5, 10, 20 or 50 topics). As topic modeling is an unsupervised method, there is no direct way to validate the discovered set of topics. However, there has been in recent years some research into computing a "coherence" value for each topic by leveraging external resources (Newmann et al. 2010; Aletras and Stevenson 2013). Based on the recommendation of Röder, Both and Hinnerburg (2015), we have decided to proceed with the normalized point-wise mutual information (NPMI) which builds upon the point wise mutual information (PMI) of any pair of words in the identified topic as defined in the following equation.

$$PMI(x, y) = log \frac{p(x, y)}{p(x)p(y)}$$

PMI is thus the log of the ratio of the observed co-occurrence frequency to the frequency expected under independence. When normalized, it can take values between -1 and 1 where -1 indicates terms never occurring together, 0 indicates a co-occurrence behavior similar to independence and 1 indicates complete co-occurrence. When aggregated at a topic level (through average across all word pairs in the topic), it provides an estimation of how likely two words in the same topic are to be found close together. We make use of the Palmetto library[3] where co-occurrence statistics are extracted from the Wikipedia corpus.

Note that we have compared the behavior of NPMI with other coherence measures by computing coherence values of 85 topics based on the 6 coherence measures analyzed in Röder, Both and Hinnerburg (2015). From those, we derived the median coherence rank of each topic based on the different metrics. Figure 6 plots that median rank on the y-axis against the NPMI value on the x-axis, showing the existence of a linear correlation between the two. This observation confirmed us in the use of NPMI which behaves similarly to other coherence measures and has a fairly interpretable definition. In particular a negative NPMI means that the

[3] https://github.com/AKSW/Palmetto/wiki/How-Palmetto-can-be-used

co-occurrence behavior is worse than what would occur if it was independent. In most of the remainder analysis, we will restrict our focus to topics that have a positive NPMI.

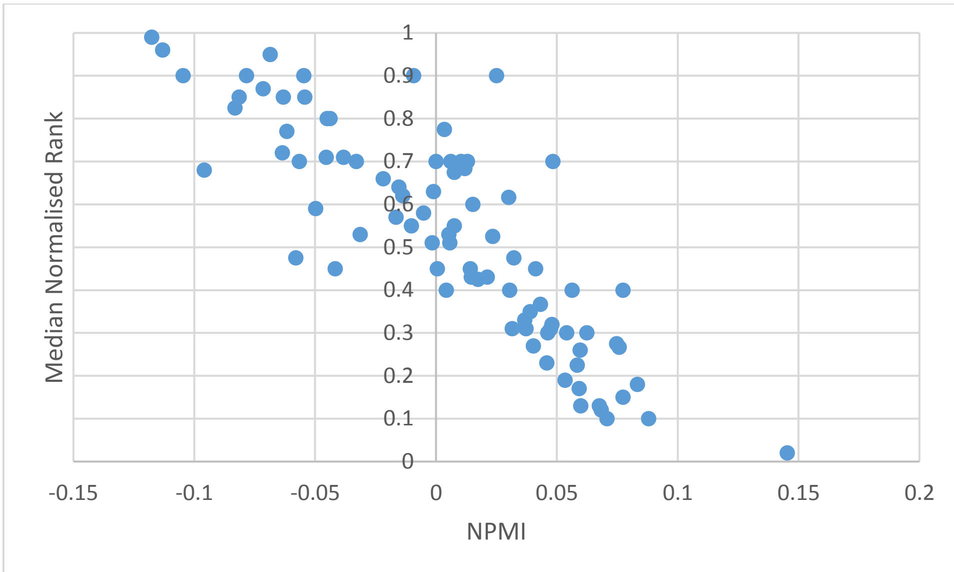

Figure 6: Scatterplot of Median Normalised Rank from a set of Coherence Metrics against NPMI measure

Results Number of Topics

Given space constraints, we cannot provide an extensive description of the topics that resulted from our experiments. However, as an example, when we run the topic model for all tweets, each one treated as an individual document, and searching for 10 topics, we obtain the topics presented in Table 4 with their coherence, likelihood and the list of the 10 most frequent words of each topic. Note that, as all tweets contain the terms “decision” and “making”, they do not appear in the lists but should be taken into consideration when seeking to associate a coherent theme to each topic. For instance, the first topic appears to be linked with opinions about decisions in sports. The second topic, slightly less obvious, would be associated with opinions on outcomes of individual and governmental choice and the third one seems related but less focused on opinions and more focused on the influence of people in social choice. As we go down the list, topics will become increasingly incomprehensible to humans. As will be the case for all experiments, there is often one or two dominating topics in terms of likelihood which corresponds to a “catch-all” category and which are seldom insightful. In Table 4, this refers to the last topic.

Table 4 : Top-Ten Words, Coherence and Associated Probability of Topics Discocvered by running LDA with all tweets treated as individual document. The topics are sorted by decreasing value of NPMI

p	npmi	words
0.013	0.05166	poor, good, game, bad, play, great, ball, time, team, terrible
0.005	0.03509	bad, people, responsible, feel, happiness, upsets, make, government, process, confidence
0.01	0.01187	women, process, people, government, power, public, local, part, involved, national
0.012	0.00024	data, business, process, analytics, management, research, improve, part, big, marketing
0.012	-0.01022	process, make, good, people, brain, skills, important, based, leadership, life
0.025	-0.02503	life, bad, make, made, good, time, wrong, people, skills, hard
0.005	-0.04611	shared, improve, creativity, video, care, skills, problem, games, focus, solving
0.003	-0.05546	life, choice, wrong, living, made, make, sleep, big, major, important
0.003	-0.05596	free, tips, skills, white, tools, insurance, street, list, squirrel, email
0.912	-0.07769	good, college, person, constantly, asshole, sarcastic, time, today, final, process

Before exploring in more details the topics that surfaced, we report on Table 5 the number of topics with positive NPMI for each configuration and each possible number of topics. We also compared configurations with and without lemmatization, i.e., with or without grouping together the inflected forms of a word so they can be analyzed as a single item. In lemmatized text, the terms “make”, “made” are replaced with their common lemma “make”. To implement the lemmatization, we rely on the lemmatizer provided with the Stanford CoreNLP pipeline (Manning et al., 2014).

As can be seen from Table 5, aggregating all tweets into one single document is detrimental to the quality of topics that are extracted. Indeed, when we ignore retweets, we never obtain any topic with positive NPMI, regardless of the number of topics set to discover and when retweets are included only 1 or 2 surfaces. Therefore, contrary to what is recommended in the literature, keeping each tweet as an individual document appears preferable. The question of whether to incorporate retweets is less obvious. In this research, in order to foster a diversity of topics, we have chosen to use the configuration that yielded a higher number of topics with positive NPMI, i.e., the one without retweets. Another justification for this choice is that retweet are only repetition of someone else's statement and as such, when focusing on understanding the meaning of the decision tweets, repeats do not contribute any information and should thus be ignored. Finally, we observe that lemmatization consistently yields a higher number of topics with positive NPMI coherence.

Table 5 : Number of topics with positive coherence for each experimental configuration. Suffix L added to the number of topics indicate lemmatised data as input.

Include Retweets	**Aggregate in one document**	**Number of topics**							
		5	5L	10	10L	20	20L	50	50L
FALSE	FALSE	3	4	4	8	7	16	15	24
TRUE	FALSE	0	2	3	6	2	9	9	20
FALSE	TRUE	0	1	0	1	0	2	0	2
TRUE	TRUE	0	1	1	1	1	2	2	3

As we proceed, our results thus correspond to an input dataset without retweet and where each tweet has been lemmatized and is considered as an individual document. Overall out of the 85 possible topics, we found 52 topics with positive NPMI and we cover through our analysis 44 (about 85%) of them. To guide the interpretation of the NPMI values reported in the following paragraphs, we provide a few statistics about their range within those 44 topics. The maximum value is 0.145 though the 95% percentile is 0.087 and the 75% 0.062. Median value is 0.045 and minimum value is 10^{-4}. We group the topics based on overlap of their 10 most frequent words. This leads to the emergence of common themes, which we present in the following paragraphs. For each theme, we provide a summary table of the words that compose it along with information about which experiment they are associated to (i.e., topic size).

Opinions on Decision Making in Sports (Table 6). The first emerging theme from our analysis is focused on decision making in sports ("*game, team, play, player*"). This theme consistently arises throughout our experiments (and often in multiple variants) as can be seen in Table 6 with a large overlap of words throughout and relatively high coherence values. We note the presence of qualifiers ("*poor, good, bad, great")* thus hinting that such tweets contain specific opinions about the quality of the decision making process. We observe some nuances in the topics, for instance, 20B seemed aligned with college level sports, 50C with baseball and 50D with referee and player decisions.

Table 6 : Summary of the Topic on *Opinions on Decision Making in Sports*.

	5	10A	10B	20A	20B	50A	50B	50C	50D
npmi	*0.056*	*0.043*	*0.046*	*0.058*	*0.017*	*0.068*	*0.053*	*0.048*	*0.059*
game	y	y	y	y	y	y	y	y	y
good	y	y	y	y	y	y	y	y	
play	y	y	y	y	y	y	y	y	
poor	y	y	y	y	y	y	y	y	y
bad	y	y		y		y	y	y	y
time	y	y	y	y			y		
great	y	y		y				y	
player	y		y	y			y		y
team	y	y			y	y			y
improve	y		y				y		
year		y			y	y			
ball				y			y	y	
skill		y		y					
coach					y	y			
Other			ability, creativity, final		college, lebron, today	terrible, win	pass	accuracy, arm, throw	captain, football, ref, referee, world

People's Decision Making Skills (Table 7). This second theme is of a more philosophical nature and appears under multiple subthemes though all related to People's Decision Making Skills. The first subtheme, which occurred consistently for each experiment size, is associated with the first four columns in Table 7 and displays an emphasis on judgment ("*bad, wrong, good, hard, feel*") on decision making for important decisions ("*life, people*"). We note that the exact same topic surfaced for the 5-topic and 20-topic experiment and that the 10-topic version only differs by one word from those and the 50-topic version by only two. Such unexpected alignment attests of the universality of the discussion about personal decision making skills, whether one's own but also other people's decision making skills. The second subtheme, shown in the last two columns of Table 7, suggests greater emphasis on emotions and reaction towards the decision making process ("*happiness, upset, responsible*"). Finally, two other topics appeared in the 20 and 50 size experiments which allude to relation between time and decision making skills with common vocabulary: "*day, today, watch, year, night*" in addition to "*good, life, bad, time, skill*" of the main theme.

Table 7 : Summary of the Topic on *People's Decision Making Skills*

	5	**10**	**20**	**50**	**20A**	**50A**
npmi	*0.077*	*0.077*	*0.077*	*0.071*	*0.039*	*0.047*
bad	y	y	y	y	y	y
feel	y	y	y	y	y	y
life	y	y	y	y	y	y
people	y	y	y	y	y	y
thing	y	y	y	y		
time	y	y	y	y	y	
wrong	y	y	y	y		y
good	y	y	y			
hard	y		y	y		
skill	y	y	y			
responsible					y	y
happiness					y	y
upset					y	y
Other		poor		change, hope	lose, money	fear, task

Patient's Decision Making (Table 8). The third recurrent theme relates to patient's decision making ("*patient, health, medical, clinical*"). While it did not occur in the 5-topic experiment, it did consistently occur in the other experiments. We note a likely link to the concept of shared decision making which promotes making treatment decisions via a discussion process between doctor(s) and patient. Observe that while "*share*" and "*shared*" could have been associated with the same lemma, the lemmatization algorithm kept them distinct most probably because of the polysemy of share and of shared which, for the former can be understood as a noun and not only as a verb and for the latter as an adjective and a verb.

Table 8 : Summary of the Topic on *Patient's Decision Making*

	10	20	50
npmi	*0.013*	*0.075*	*0.083*
care	y	y	y
health	y	y	y
patient	y	y	y
share	y	y	y
shared	y	y	y
clinical		y	y
medical		y	y
support	y		y
Other	part, problem, process, solve	brain, improve, study	healthcare, tool

Decision Making and Management. (Table 9) This theme, present in all but one experiment size, discusses the relationship between decision making and leadership and management skills. There is a nuance of development and/or quality judgment through the presence of words such as "*improve, learn, skill, good, great*".

Table 9 : Summary of the Topic on *Decision Making and Management*

	5	20	50A	50B
npmi	*0.049*	*0.062*	*0.037*	*0.06*
skill	y	y	y	y
management		y	y	y
business	y	y		
improve	y	y		y
leadership	y	y	y	
process	y	y	y	
problem	y			y
good		y	y	
learn		y	y	
part		y		y
great			y	y
other	brain, change, choice, life	success	leader, team, work	characteristics, solve, strategy, training

Decision Making and Public Decisions (Table 10). The fifth prevalent theme across our configurations deals with people's participation in public decision processes. Note that this topic does not imply opinions but rather statements. Versions 5, 10, 20A and 50A indicate a specific focus on women's involvement and youth involvement (for 20A and 50A). Versions 20B and 50B present a focus on government decisions by lawmakers rather than by the public *("government, vote, bill, law").*

Table 10 : Summary of the Topic on *Decision Making and Public Decisions.*

	5	**10**	**20A**	**20B**	**50A**	**50B**
npmi	*0.025*	*0.054*	*0.041*	*0.088*	*0.037*	*0.068*
process	y	y	y	y	y	y
power	y	y	y	y	y	y
woman	y	y	y		y	
people	y	y	y		y	
involve	y	y	y		y	
government	y	y		y		y
local		y	y			y
youth			y		y	
young			y		y	
vote				y		y
public		y				y
political		y		y		
policy		y				y
participation			y		y	
level			y		y	
Other	body, person, good, give			supreme, president, obama, law, court	role	support, council, bill

One related set of topics (surfacing in all experiments but the one with 5 topics and reported in Table 11) also deals with public policy, specifically with the National Action Charter from Bahrain which, incidentally, celebrated its 15th anniversary in February 2016[4].

Table 11 : Summary of the Topic on *Bahrain National Action Charter*

	10	20	50
npmi	*0.012*	*10^{-4}*	*0.0146*
action	y	y	y
bahrain	y	y	y
citizen	y	y	y
national	y	y	y
participate	y	y	y
confidence	y	y	
forward	y	y	
move	y	y	
legislation	y		y
	decide	cross, skill	charter, feb, nationalactioncharter, worried

Other themes did not occur as consistently in the analyses but were present in at least two experiments and seemed to point to a central concept[5]. One of them deals with the relationship between Sleep and Decision Making with common words such as "*deserve, hour, important, life, sleep, trouble"* hinting either at trouble sleeping or getting a good night sleep for improved decision making. Another stream of tweets appears related to Life Changing Decisions through vocabulary "*change, choice, life, live, major, prove*".

[4] We have not investigated the timing of the associated tweets so we cannot draw a formal link between the occurrence of the topic and the anniversary.
[5] Two other themes occurring in two topics were more ambiguous. The first one is associated with common words "*free, home, list, tip, tool, white*" which evokes decision making advice and the second one is associated with "*base, good, power, people, poor, process*", certainly less defined than others.

Finally, we mention two topics that only appeared in the 50-topic configuration because they are associated with high NPMI value. The first one pertains to Decision Making and Education ("*child, education, involve, parent, process, school, skill, society, student*") and the second one with Behavioral Decision Making ("*Bias, brain, cognitive, emotion, human, influence, process, psychology, research"*) a topic more familiar to decision making researchers than most that have been mentioned so far and which corresponds to the highest NPMI value.

Topics for specific subsets

In this section, we compare topics for tweets associated with specific characteristics, as we observed in Section 3 the relevance of gender, polarity and also day of the week on decision tweet characteristics. For the following analyses, we have chosen not to include retweets, to consider each tweet as an individual document and to run the topic models with N= 10 topics and based on a lemmatized version of the tweets. For each of the dimensions of interest, we assemble separate sets of input documents for each possible values of the variables and we subsequently run the LDA algorithm on each such dataset.

We report in Table 12 the NPMI values associated with the topics aligned with each of the major and minor themes that were uncovered in the general analysis. This summary table enables to identify themes that are fairly universal such as Opinions on Decision Making in Sports, People's Decision Making Skills and Decision Making and Public Decisions and to a lesser extent, Decision Making and Management. We note that Patient's Decision Making, one of the five main themes identified, only surfaced for unspecified gender and for positive polarity. Finally we observe that two side themes, Sleep and Decision Making and Decision Making and Education, are absent from all experiments. The unexpected theme about Bahrain National

Action Charter appears within the neutral comments and male authors subsets. Similarly, tweets about Life Changing Decisions surface within the Fridays, negative polarity and male authors' subsets.

Our analysis of topics for subsets also led to the following observations. First, despite our previous comment on catch-all topics, i.e., the ones with high likelihood but typically low coherence, we note that for tweets authors reporting their gender as male, the most likely topic is People's Decision Making Skills (58% of the "male" tweets) with a relative high NPMI value around 0.06. This theme also shows as the prevalent topic for tweets classified as ambivalent (around 95% of such tweets) or neutral (96%) and for Tuesdays, Thursdays and Fridays (respectively 88%, 94% and 60%). On Sundays, the most likely topic belongs to the Decision making in Sports theme with a likelihood of 47%. Second, some analyses led to the emergence of hybrid topics. For instance, for Wednesdays tweets, we found a hybrid topic between Decision Making and Management and Behavioral Decision Making Research (process, business, problem, brain, affect, good, bias, skill, leadership, solve).

Table 12 : NPMI values (coherence) for each subset of tweets and each theme.
Note: Cells with two entries indicate that two topics were found in the corresponding experiment

	Opinions on Decision Making in Sports	*People's Decision Making Skills*	*Patient's Decision Making*	*Decision Making and Management*	*Decision Making and Public Decisions*	*Sleep and Decision Making*	*Life Changing Decisions*	*Decision Making and Education*	*Behavioral Decision Making*	*Bahrain National Action Charter*
Female	0.051	0.091		0.06 & 0.019						
Male	0.059	0.039		0.031	0.031		0.064			0.005
Unknown	0.06	0.076	0.052	0.061						
Ambivalent	0.065	0.097								
Positive	0.075		0.014		0.023					
Neutral	0.057	0.038			0.046					0.017
Negative	0.062 & 0.047				0.022		0.033			
Monday	0.043	0.077		0.014	0.029					
Tuesday	0.065	0.074 & 0.043		0.01	0.034					
Wednesday	0.061	0.078		0.06	0.028					
Thursday		0.065 & 0.028		0.043	0.041				0.066	
Friday	0.06	0.076			0.033		0.033			
Saturday	0.065	0.07			0.006					
Sunday	0.058	0.065			0.02					

6. Discussion and Conclusion

In this paper, we provide a descriptive analysis of Twitter data focused on decision making. We systematically reviewed the relationship of gender, polarity, day of week, retweet and presence of “Decision making” on the tweeting behavior and compared to ordinary tweets. We proceeded to focus on the content of tweets by applying topic modeling to the tweets’ body. In that effort, we identified five main recurring themes: (i) Opinions on Decision Making in Sports, (ii) People’s Decision Making Skills, (ii) Patient’s Decision Making, (iv) Decision Making and Management and (v) Decision Making and Public Decisions. Five minor themes also emerged with some consistency: Sleep and Decision Making, Life Changing Decisions, Decision Making and Education, Behavioral Decision Making and Bahrain National Action Charter.

Taken individually some of the themes may be surprising. Overall, they reflect for the most part familiar topics in Decision Sciences broadly defined. Several themes align naturally with prescriptive and descriptive Decision Analysis at a general level: People’s Decision Making Skills and Decision Making and Management but also the minor themes such as Life Changing Decisions, Sleep and Decision Making and possibly Behavioral Decision Making. Then, the domain specific themes are common in Operations Research if not Decision Analysis where Education, Healthcare and Sports have long been studied. Finally, the themes around public policy making and public decisions are aligned with more Economics-driven perspectives on decision making. At a high level, decision making tweets, or maybe some chosen subsets, are pertinent to existing research in our domain though their content appears to be more strongly associated with appreciation or feelings about facts rather than facts themselves. As we pointed out for People's Decision Making Skills for instance, though this happens in most of our themes, there is a strong presence of qualifiers and sentiment words. The statistical analysis did also reveal that decision tweets are more opinionated that ordinary tweets. In that context, tweets appears more naturally suited to research that focuses on behavioral aspects of decision making, for instance the study of the expression of regret.

One of the recurring observation throughout the analyses is the influence of gender. Decision tweets are less likely to be written by female than ordinary tweets and the polarity distribution of those tweets differ significantly according to gender with male being more often positive. Furthermore, one of the main theme has a specific reference to women, specifically related to the role of women in public policy decisions (whether encouragement or observation, this has not been determined). The life changing decisions theme surfaced only within the male subset while at the same time, People's Decision Making Skills topic is a prevalent topic for male authors but not for female or unknown. While there already is ample research on the influence of gender on risk aversion, managerial decision making, consumption decisions, ethical decision making, we believe that the set of findings that emerges here broadens the discussion. Rather than link gender to the preferences of decision makers, those findings raise the question of whether male and female have a different interest in decision making, a different way of sharing and expressing their internal decision process, a different way of evaluating other people's decision making. We believe that this could constitute a worthy domain of investigation through Twitter and beyond.

Another prevalent focus is sports. It appeared when we studied the source of the outlier days in terms of decision tweet volume, arose as one of the main topics in our general analysis with multiple variants and also represents the most likely topic associated with Sundays tweets besides being a key topic almost every day of the week. Sports and decision making appear deeply connected in people's mind and a frequent source of shared observations. This raises the question of whether sports could be further leveraged in Decision Analysis, among others to educate people in terms of common decision making concepts (e.g., preferences, rationality) and errors (decision versus outcome). How could Twitter be actively used for such endeavors?

Finally, the concept of time appears in several places in our study. In particular, the day of week has an influence on volume and polarity and one of the subthemes around Personal Decision Making focuses specifically on time through vocabulary "*day, today, watch, year, night*". The minor topic on Sleep and Decision Making could also be understood as relating to time. Those findings may also warrant further investigation, among others as to the relationship of our occupation (through day of week proxy and night/day) to our interest and focus on decision making.

Naturally, as we relied on fairly simple approaches, whether for the statistical or text mining parts, further work in validating and interpreting the statistical findings would be justified. In particular, the datasets that we have assembled could be further leveraged as follows. Inference could be used to obtain a richer description of the tweet authors, specifically through algorithms designed to detect gender, age, and location from the tweet body, the latter to study cultural specificities. Second, the natural language processing could be expanded. We intend to make use of named-entity recognition services and possibly relation extraction algorithms to extract further knowledge from the tweets. We will explore more semantic-based approach which will enable us to go beyond the lexical representation of words and group together those with similar ore related meaning (for instance through word2vec previously mentioned). Finally, and more importantly, we plan to focus on detecting and analyzing more Decision Analysis specific concepts such as expressions of trade-offs or biases. Twitter could also be used to discover common decision situations (college application for instance) and eventually be leveraged to generate more creative alternatives, better understand the associated uncertainties and better understand one owns preference dimensions by understanding what criteria other have considered. Such efforts, however, requires a more sophisticated retrieval and processing of text.

Altogether, we believe that Twitter discussions are relevant to Decision Analysis research, more obviously suited for behavioral investigations than normative decision support. While promising, leveraging tweets for specific research questions, by opposition to the all-encompassing high-level study that is presented in this paper, will require a significant level of information retrieval and natural language processing expertise.

As a clarification, our objective in this research was to make use of the information that is spontaneously shared via the Twitter platform. There are research opportunities as well in more active experiments where Twitter is actively used to elicit reactions from users on a given question or task by prompting reactions through tweets posted by researchers. This type of research is beyond the scope of the current effort.

7. References

Aletras, N. and Stevenson, M., 2013. Evaluating topic coherence using distributional semantics. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)–Long Papers* (pp. 13-22).

Blei, D.M., Ng, A.Y. and Jordan, M.I., 2003. Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), pp.993-1022.

Golder S, Macy M. Social Science with Social Media. ASA Footnotes, 2012;40(1).

Hong, L. and Davison, B.D., 2010, July. Empirical study of topic modeling in Twitter. In *Proceedings of the first workshop on social media analytics* (pp. 80-88). ACM.

Hong, L. and Davison, B.D., 2010, July. Empirical study of topic modeling in Twitter. In *Proceedings of the first workshop on social media analytics* (pp. 80-88). ACM.

Li, C & Liu, Y 2015, Improving named entity recognition in tweets via detecting non-standard words. in *ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Proceedings of the Conference.* vol. 1, Association for Computational Linguistics (ACL), pp. 929-938,

Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55-60.

McCallum, Andrew Kachites. "MALLET: A Machine Learning for Language Toolkit." http://mallet.cs.umass.edu. 2002.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S. and Dean, J., 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119).

Newman, D., Lau, J.H., Grieser, K. and Baldwin, T., 2010, June. Automatic evaluation of topic coherence. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (pp. 100-108). Association for Computational Linguistics.

Princeton University "About WordNet." WordNet. Princeton University. 2010. <http://wordnet.princeton.edu>

Ramage, D., Dumais, S.T. and Liebling, D.J., 2010. Characterizing Microblogs with Topic Models. *ICWSM*, *10*, pp.1-1.

Röder, M., Both, A. and Hinneburg, A., 2015, February. Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining* (pp. 399-408). ACM.

Vosecky, J., Jiang, D., Leung, K.W.T. and Ng, W., 2013, October. Dynamic multi-faceted topic discovery in Twitter. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management* (pp. 879-884). ACM.

Zhao, W.X., Jiang, J., Weng, J., He, J., Lim, E.P., Yan, H. and Li, X., 2011, April. Comparing Twitter and traditional media using topic models. In *European Conference on Information Retrieval* (pp. 338-349). Springer Berlin Heidelberg.