Tweeting about health technologies during the *Movember* campaign: Who is involved and how do they frame their discourse?

An exploratory study investigating the identity of Twitter users tweeting about technology and treatments of Prostate and Testicular cancers, and how they frame their discourse

Author: Balsam Awlad-Wadair
Supervisors: Dr.Michel Ehrenhard and Anna Priante
External examiner: Dr.Ir. Fokko Jan Dijksterhuis
NIKOS Research Group
Advanced Technology Bachelor Thesis
June, 2016

Abstract
Cancer awareness campaigns like *Movember* have a great potential of effectively educating people about cancer and ways of its early detection. In the past few years, the *Movember* campaign relied on social media platforms such as Twitter to reach its objective of raising awareness about men's health, in particular Prostate and Testicular cancers early detection and treatments. However, little is known on whether such online campaigns promote people's knowledge and awareness about technologies and treatments for cancer. This article presents an analysis of the United Kingdom (UK) 2014 *Movember* campaign on Twitter and investigates 1) whether Twitter users are tweeting about technologies used for the detection and treatments of Prostate and Testicular cancers; 2) and how these users frame such discourse online. Furthermore, social identity theory is used to identify the profiles of those tweeting about technologies and treatments. The following study provides three main contributions to existing literature. First, the findings suggest that only 0.5% of the tweets from the UK 2014 *Movember* campaign are related to Prostate and Testicular cancers technologies and treatments. Second, results obtained from sentiment and semantic analysis show that Twitter users adopt mostly negative tone in their tweets and they frame their discourse around two main thematic clusters: generic treatments (such as *Radiotherapy, Chemotherapy, Immunotherapy*) and specific diagnostic techniques (such as *Transrectal Ultrasound* and *Proton Beam Therapy*). Third, people tweeting about technologies and treatments were predominantly classified into the 'Occupation' social identity category. These findings offer basis for further research on the topic and provide practical implications for campaign organizations.

Keywords
Prostate Cancer — Testicular Cancer — Movember — Social Media — Health Communication — Social Identity — Technology

Contents

Introduction

Multiple campaigns with a mutual aim of raising awareness about different types of cancer and its early detection are held annually on different social media platforms. Twitter (an online micro-blogging service) is a solid example of a platform which hosts cancer early detection campaigns, such as *Mamming* (breast cancer), *Movember* (prostate cancer), *DaveeDay* (pancreatic cancer) and *HPVReport* (cervical cancer) (UTwente, 2014). For example, the growing annual campaign of *Movember*, which aims to raise awareness about men's health, in particular prostate and testicular cancer, relies heavily on social media platforms like Twitter. The primary goal of the campaign is to increase early detection, diagnosis and effective treatment which can be made possible by the use of health technologies. Although the outreach of such campaigns on Twitter is large (Rosemary Thackeray & Draper, 2013), the question of the effectiveness of such campaigns on providing awareness and encouraging early detection of the disease remains ambiguous (Bravo & Hoffman-Goetz, 2015b; Prasetyo et al., 2015). In particular, little is known on whether such online campaigns promote people's knowledge and awareness about technologies and treatments of cancer.

Although advancement in technologies aid early detection and treatment of cancer, as well as health research in general (Lang, 2011), the awareness of such technologies must be improved in order for it to function as desired (Tait, 2015) and therefore assist in early detection and treatment of cancer. Many people remain unaware of technologies used for cancer treatment and early detection (Movember Foundation, 2014), which is a drawback and is now an objective of many cancer awareness campaigns.

Using data from the United Kingdom (one of the largest English speaking countries where the *Movember* campaign runs) 2014 *Movember* campaign on Twitter, this research aims to provide an exploratory analysis on whether people showcase knowledge of health technologies and treatments and how they frame such discourse during the campaign. Sentiment and semantic analysis of *Movember* tweets is vital to determine whether Twitter users talk about technologies used for treatment and early detection of prostate and testicular cancers. Furthermore, this research investigates who are the people (based on social identity theory) which are more likely to tweet about health technology. This linkage of identity and behavior can be made as social identity can be used to provide explanations of people's attitudes and/or behavior (Scott, 2007). This will be elaborated in more detail later in this paper.

More specifically, the following research questions need to be addressed:

1. Do Twitter users participating in the UK 2014 Movember campaign tweet about health technologies and treatments of Prostate and Testicular cancers?
2. What are the main technologies and treatments mentioned by these users?
3. How do these users frame their tweets about technologies and treatments of Prostate and Testicular cancers?
4. In the case of participants tweeting about health technologies, to which social identity category do such users belong to? In other words: Who tweets about health technologies and treatments related to prostate and testicular cancers?

The result of this study will therefore be valuable to organizers of the *Movember* campaign, as it shows which Twitter users participating in the *Movember* campaign are knowledgeable of technologies and cancer treatments, which aids the motives of the campaign (Bravo & Hoffman-Goetz, 2015b). This result will help determining whether the campaign is effective in raising awareness of technologies used for treatments and early detection of cancers, specifically Prostate and Testicular cancers online. Likewise, this research aims at deepening our knowledge about how such campaigns develop online, what type of users are involved, what content do they distribute and most importantly hints about the campaign's effectiveness.

This thesis is organized as follows. First the literature review and theoretical framework are presented. Second, the methodological approaches used for answering each of the research questions are described. Third, the results for this research are given with an analysis and brief discussion. The paper then ends with a conclusion and suggestions for further research and practical implications.

1. Literature Review

Studies made previously revealed that social media platforms (such as Twitter, Facebook, Pinterest, etc) have altered the way people communicate about health information. An example portraying this influence is seen from how individuals are now seeking what their peers are saying online about health and treatments before accepting a doctor's advice (Ratzan, 2011). Furthermore, it has been stated in literature that this is most frequently done by young well-educated adults (Percheski & Hargittai, 2011). Twitter in particular is considered a primary platform enabling information dissemination (Pew Research Center, 2014). Subsequently, the examination of the content of *Movember* related social media publications such as tweets becomes important to evaluate whether conversations involve health related topics like technologies and treatments used.

Limited research has been made on Twitter conversations related to the *Movember* campaign and their content. One study conducted in relation to the 2013 *Movember* campaign

targeting Twitter users in the United States, Canada and the United Kingdom has shown that few tweets ($n = 84$, 0.7% of 12,666 tweets) were health-related and had the potential of raising awareness about men's health and prostate cancer (Bravo & Hoffman-Goetz, 2015a). Similarly, another study which focused solely on the 2013 *Movember* tweets from Canadian Twitter users revealed that there were significantly less health-related ($n = 673$) than nonhealth-related ($n = 3549$) tweets. The study also looked in detail at the words used during twitter conversations of the Movember campaign and exhibited that few tweets (0.6% of all tweets) referenced prostate or testicular cancers (Bravo & Hoffman-Goetz, 2015d). Likewise, Jacobson & Mascaro (2016) found through their research on the *Movember* campaign Twitter conversations that limited true conversation is taking place, regardless the given objective of the campaign which is to facilitate conversation about men's health including Prostate and Testicular cancers.

In regard to the correlation between social identity of individuals and their knowledge of technology and/or cancer treatments, there is a lack of existing literature. It has been reported by Lee et al. (2006) that self-identity has a significant direct and indirect influence on technology acceptance. This acceptance counts as a behavior and attitude towards technology, which can be predicted by the social identity of individuals as it can give explanations of their behavior (Scott, 2007). In order to elaborate further on the relationship between social identity of individuals and the likelihood to showcase knowledge of technology and treatments in online cancer awareness campaigns, we chose to investigate the identity of Twitter users tweeting about technology and treatment. This will make contributions to existing literature, as little to no research has been made regarding this topic.

2. Theoretical Framework

Based on the available literature examined as preparation for this research, it can be seen that tweets related to health and treatments of prostate and testicular cancers hold a relatively low percentage from the overall numbers of tweets from the *Movember* campaign. Therefore, it can be hypothesized that there will be relatively few cases where Twitter users will mention technologies or treatments used for Prostate and Testicular cancers. On the other hand, as studies have shown the majority of tweets are nonhealth-related. In cases where people do mention terms related to the technologies and treatments used for Prostate and Testicular cancers, it may be the case that it is mentioned for promotional purposes by organizations (Bravo & Hoffman-Goetz, 2015c) or by professionals working in related fields. Therefore, looking at the identity of people tweeting about technology can inform better about the users who have the potential to spread discourse on health technologies.

A link between Twitter profiles of users mentioning technologies and treatments is also made to social identity, which serves as a theoretical choice to operationalize the third research question regarding 'who' is tweeting about technologies and treatments. The motivation of creating such linkage revolves around the language used by each user in defining who they are on their Twitter profile summaries, as the language used by individuals is considered to have a strong association to their social identity (Tamburrini et al., 2015). Similarly, grouping users will possibly give an explanation to their behavior (Scott, 2007) which is in this case whether they tweet about health technology and treatments.

This paper tries to find out whether people tweeting about the UK *Movember* campaign on Twitter are showcasing knowledge of technologies and treatments used for prostate and testicular cancer through their tweets or conversations on Twitter. In the absence of a large content of published research on linking social identity and twitter conversations, it is believed that a preliminary study is critical in order to

- Understand if there is a connection between identity and conversations in order to find out whether people who tweet about technologies and treatments belong to a specific social identity category
- Develop methods of classifying twitter profiles into different social identity categories
- Develop methods to identify and analyze conversations that can be linked to technologies and treatment of prostate and testicular cancer
- Foresee potential research implications or challenges

2.1 Social Identity Classification

This paper uses a social identity classification developed by Priante et al. (2016). The aim of using such classification is to categorize the people who tweet about technology and treatments related to prostate and testicular cancers. This classification combines both Identity Theory (IT) and Social Identity Theory (SIT) and is a re-elaboration of the works of Deaux et al. (1995); Ashforth et al. (2008, 2016).

To develop further understanding of the basis of social identity classification, the definitions of the theories involved in the formulation of the classification have been studied. Based on the work of Stryker & Burke (2000) identity is composed of meanings that individuals appoint to the different roles they normally play in highly differentiated contemporary societies. Identity theory (IT) suggests that identity of individuals is established in social groups/network of relationships/roles, and therefore people tend to attach their identity to a role they play in those different groups. On the other hand, Social Identity Theory (SIT) was developed by Tajfel and Turner in 1979, and it states that people have the tendency to formulate their identity according to an individual's sense of who they are based on group membership(s) (Tajfel & Turner, 1979).

The social identity classification used in this study consists of 5 categories:

- **Relational**: self-definitions based on relationships and social roles played by the individual. Examples include: father, mother, lover, fan of someone

Table 1. Social Identity category and representative Twitter profiles' descriptions

Category	Representative profile description
Relation	*"Saw One Direction and 5sos live on the 25/10/13 - best night of my life!! 0/9"*
	"Love my son and my chihuahua&chorkie! xxxx"
Occupation	*"Stand up comedian. Broadcaster. I think. I write. Sometimes I should probably put a step in between those two."*
	"Resident DJ @ Air Nightclub, Producer, 21"
Political	*"voice for every single cause to STOP cruelty"*
Ethnic/Religion	*"Love Jesus but i give him hell"*
	"young, gifted & black"
Stigma	*"Just an Anon... Sucidal, deression, self harm. — They said she was an attention seeker — Female 15"*
	"Sports geek. Music geek. Straight edge vegetarian."
Not Applicable	*"Melbourne based environmentally focused clothing individual designs on an extremrly limited run! EST. 2013"*
	"Always find a reason to be happy. Smile. Its a Sunnah. =) "

- **Occupational**: those defining themselves according to their occupation or profession. It also reflect individual vocations/avocations
- **Political**: identities which show political affiliations and parties, as well as being part/enthusiastic about social movements and collectives
- **Ethnic/Religious**: self-definitions based on memberships of ethnic/religious groups. This category also includes those expressing rejection of religion (e.g. 'Godless')
- **Stigma**: refers to identities which belong to a stigmatized group. A stigma is a negative label attached to people who are considered different from what is considered the normal according to the society and its cultural norms. Examples of typically stigmatized groups of people are homosexuals, alcoholics, and overweight people.

3. Methods

This section presents the methodological approach of this research. First, the preparation of the data-set used for the study is explained. Second, the methods employed to answer each research question are described.

3.1 Preparing the data-set

Data from the Twitter data grant (a project initiated by Twitter with the aim of providing research institutions access to Twitter's public and historical data (Krikorian, 2014)) is used for this paper. The set of tweets used for this study were all published by Twitter users from the UK during the 2014 *Movember* campaign and all shared the criteria of having the hash-tag *#Movember*. The tweets were all available in English. The geographical location was collected from the profiles of the users, and the country classifier made by van der Veen et al. (2015) was used to select users only in the UK. This study focused on analyzing the tweets published by the users during the *Movember* campaign in order to identify themes and conversations about technologies and treatments related to Prostate and Testicular cancers. Additionally, the user name and profile description were used to identify the social identity of the user according to the classification discussed on section 2.1. Furthermore, blank profiles were removed from the data set as they provide no relevant information regarding social identity. In total, the data-set of this study consists of 855108 tweets sent by 326234 users.

3.2 Identifying Technologies and Treatments

In order to answer the first two research questions listed in the introduction, existing research on prostate and testicular cancers technologies and treatments provided guidance in recognizing potential themes and concepts in the tweets. A list of keywords related to technologies and treatments of prostate and testicular cancers was then made to assist this research (see section A on the appendix). With the help of basic software such as Microsoft Excel, the list of keywords was used to select the tweets about technologies.

3.3 Analyzing the content of tweets

Providing answers to the second and third research questions requires a deeper analysis of the content of tweets. In order to do this, several methods have been adopted by using the open source software NodeXL, an Excel plugin that provides tools for analysis of content, sentiment and network features.

First, the content of tweets was analyzed by using the list of keywords presented in section 3.2. Basic descriptive statistics (frequencies) are calculated to identify the list of keywords present within the data-set of tweets, as well as to determine the number of tweets containing technology and treatment related topics. Additionally, the researcher read the tweets attentively to recognize tweets which can be labeled as tweets mentioning technologies and treatments of prostate and testicular cancers.

Second, sentiment analysis is used to determine the overall tone of tweets about technology and treatment. The sentiment score is calculated from an equation (Younis, 2015) based on sets of positive and negative words proposed by Hu & Liu (2004); Liu et al. (2005)

$$Sentiment score = \sum Positive words - \sum Negative words \quad (1)$$

Third, semantic network analysis is used to analyze more in depth the content of tweets and to understand how Twitter

users frame the discourse on technologies and treatments for Prostate and Testicular cancers. Semantic network analysis is a technique to analyze text data using network metrics and visualizations (Drieger, 2013). Starting from the text of tweets, a semantic co – occurrence network of the keywords related to technology and treatment was created using NodeXL. The semantic network is made of a set of nodes (one for each keyword) and edges, which connect keywords that co-appear in the same tweet. Keywords that appear alone in the tweets appear as isolated nodes and have been excluded from the analysis. Centrality metrics (Freeman, 1979) are then calculated with NodelXL to identify the most prominent keywords within the network. Finally, cluster analysis is used to identify sets of keywords that are often used together and represent the thematic semantic cores of the online discussion.

3.4 Social Identity classifications

The last research question (question 4 listed in the introduction) aims to identify who are the people who tweet about technology. We used the social identity classification described in section 2.1 to manually annotate 1439 Twitter profiles. Two annotators were involved in this process. Several training sessions were implemented in order to derive a set of rules to assist decision making for the social identity annotations. It was decided to focus only on the summaries of the profiles provided by the users. Extraneous content which may include images, links and promotional content were not used. Due to the fact that social identity definition is to be applied only to individuals, profiles belonging to organizations, groups of individuals, fan-pages or collectivities were placed into an extra category named 'Not Applicable'. Likewise, cases where profile descriptions contained ambiguous content or value statements (quotations, religious citations, etc) which cannot be annotated were also classified as 'Not Applicable'.

Examples of representative profile summaries for each category can be found on table 1.

3.4.1 Reliability measurement

In content analysis, researchers have to account for the reproducibility of their results and thus a reliability measurement is needed in this case to determine if the social identity classification system (see section 2.1) is appropriate to be used for measuring the identity of the Twitter users. We calculated the inter-rater reliability on 300 profiles using Krippendorff's alpha (KALPHA) (Krippendorff, 2004). The reliability scores for each category were all satisfactory (higher than 0.8). Table 2 shows the exact reliability scores calculated for each social identity category.

Table 2. Table of inter-rater reliability scores of each social identity category

Category	KALPHA
Relation	0.9019
Occupation	0.8911
Political	0.9197
Ethnic/Religion	0.8910
Stigma	0.8531
Not Applicable	0.8268

4. Results and Discussion

This section presents the findings of this study, and thus provides answers to our research questions.

4.1 Basic Descriptive Statistics

Our first research question concerns whether Twitter users participating on the *Movember* campaign tweet about technologies and treatments of prostate and testicular cancers. Counting and frequencies were used to examine the number of tweets mentioning technology and treatments using the list of words created in section 3.2.

A total of 3956 tweets from the UK *Movember* data-set mentioned terms related to prostate and testicular cancers technology and treatment, which corresponds to approximately 0.5% of the overall data-set of tweets ($n = 855109$). The results obtained show almost no difference to prior studies made for the *Movember* tweets mentioned in section 1, and comes as confirmation that the discourse on Twitter from the *Movember* campaign contain little health-related information.

The second research question aims to identify the main technologies and treatments mentioned. Table 3 displays the frequencies of keywords mentioned in the UK 2014 *Movember* tweets. From the table, it can be seen that the 5 words with the highest frequency: *Chemotherapy, Immunotherapy, Genetics, MRI/MRI-scan(ner)* and *Biopsy*. This result can be expected as those terms correspond to the most common technologies and treatments correlated to Prostate and Testicular cancers (http://www.cancer.org, 2016a,b).

4.2 Sentiment and Semantic analysis

In order to answer to the third research question on how Twitter users frame their discourse about technologies and treatments, sentiment and semantic analyses are conducted. First, we looked at the overall sentiment of the discourse. Second, we investigate the importance of keywords on technologies and treatments and the presence of thematic cores in the discussion.

4.2.1 Sentiment analysis

The tweets about technologies and treatments of Prostate and Testicular cancers are deemed to have a negative tone, according to a sentiment score explained on section 3.3. The sentiment score obtained for the set of tweets was -7 and

Table 3. Prostate and Testicular cancers Technology and Treatment keywords present in tweets about the 2014 UK *Movember* campaign and their frequency of occurence

Keyword	Frequency
Chemotherapy	1208
Radiation Therapy	794
Immunotherapy	686
Genetics	455
Biopsy	342
MRI	201
Nanotechnology	103
Hormone Therapy	69
Vaccine treatment	50
Brachytherapy	47
Gleason	20
ConfirmMDx	20
Prostate-specific Antigen (PSA)	19
Radical Prostatectomy	18
Multiparametric MRI	13
Intensity Modulated Radiation Therapy (IMRT)	10
Digital rectal exam (DRE)	8
Positron-emission Tomography (PET)	8
Prostate cancer antigen 3 (PCA3)	6
4KScore	6
Cryotherapy	6
TMPRSS2:ERG	3
Transrectal Ultrasound (TRUS)	1

*Total number of tweets with keywords related to technologies and treatments was 3956

therefore has a negative polarity. This means that a negative tone is predominant within the tweets. This is expected because the list of negative words contains words like 'cancer', 'cancerous' and 'illness' which are often seen within the data-set of tweets.

4.2.2 Semantic Analysis

We look at the overall semantic co-occurrence network in order to identify how Twitter users frame the discourse of technology and treatment. Table 4 displays basic features of the co-occurrence network.

Table 4. Semantic network overall metrics

Graph metric	Value
Vertices	29
Unique edges	51
Total Edges	51
Density	0.12
Connected components	2
Maximum vertices in a connected component	27
Maximum edges in a connected component	50

There are 29 keywords (vertices) related to technology and treatments of Prostate and Testicular cancers. In total, there are 51 edges which correspond to keyword co-occurrences: one edge links two distinct keywords that are found on the same tweet. We operationalize the presence of an edge between 2 keywords as a symbol of awareness of the people who tweet about the technologies in the tweet, as they have sufficient knowledge to make such linkage.

The density (which is the ratio of actual connections between nodes within a network and all possible connections (Hansen et al., 2009)) informs about the structure of the network. It indicates that the network is quite sparse, as only 12% of possible ties are activated. This means that users usually tend to tweet about different technologies but rarely combined them in the same tweet. This can be sign of a quite fragmented discourse, where users talk about one specific technology in each tweet. Although people talk about various types of technologies, looking at the number of connected components shows an overall convergence of the discourse.

In network analysis, a connected component refers to the number of sets of vertices that are connected to each other but not to the rest of the graph. In other words, the more components, the more fragmented is the discourse. However, in our semantic network, there are only 2 connected components. The main one holds approximately 93% of the total keywords (27 out of the 29 vertices) and 98% (50 out of the total 51 edges) of the relations. It means that, despite that keywords mostly appear disconnected from each other (low density), they all together converge in one main component and thus contribute in making the discourse about technologies and treatments of prostate and testicular cancers.

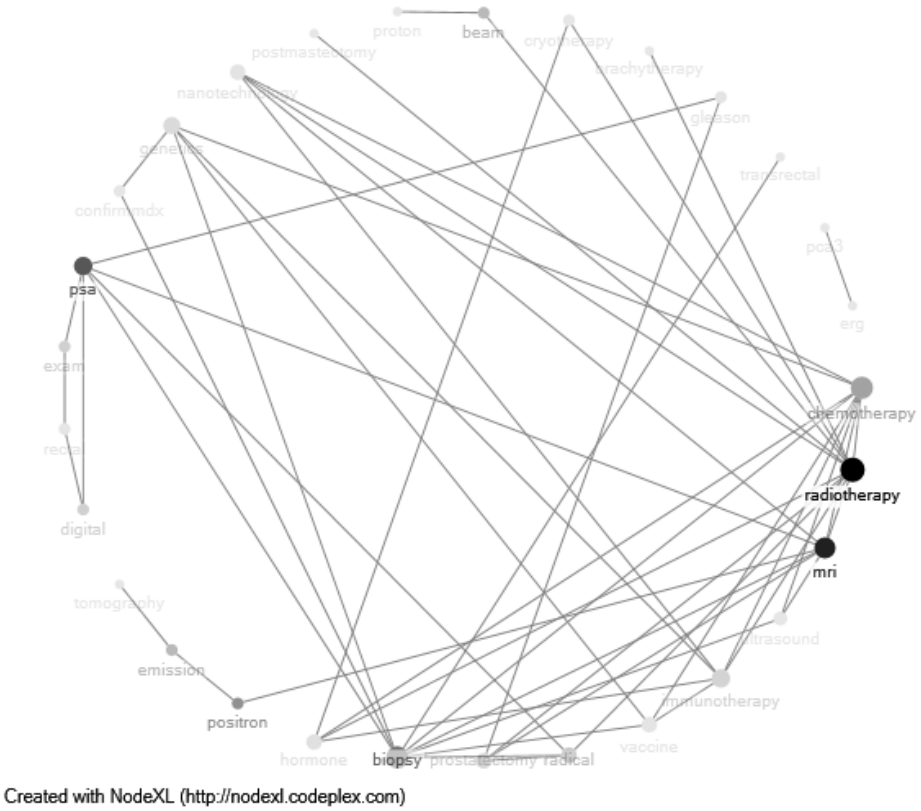

Figure 1. Semantic network of keywords. Note that node sizes are proportional to the *degree*, while the color corresponds to the *betweenness* where *high betweenness* is black and *low betweenness* is grey.

This convergence is made possible by certain keywords which are responsible for linking important parts of the network together. In order to find these important keywords within the tweets, we calculated the centrality indexes: *degree* and *betweeness* (Freeman, 1979).

In network analysis, the degree is the number of edges that are connected to a node. We refer to the degree because our semantic network is an undirected graph since keyword co-occurence implies no direction. The higher is the degree, the more important is the keyword. This means that keywords with high degree are the ones mostly used together with others in the tweets (Freeman, 1979; Wasserman & Faust, 1994). Such keywords symbolize the technologies or treatments that people are most aware of in the discourse.

In our semantic network, the degree distribution (see figure 2) shows a decrease as most of the keywords have a low degree value, while a few have a high degree. In order to overcome the issue of this variability of distribution, a cut-off value[1] must be used in order to determine the central keywords. This makes it possible to identify which one of the keywords are most prominently used within the semantic network.

[1]The cut-off value is calculated by summing the *mean* and *standard deviation* of all the *degree* values. The same process is also used to calculate the cut-off value for the *betweenness* (Pavan, 2012)

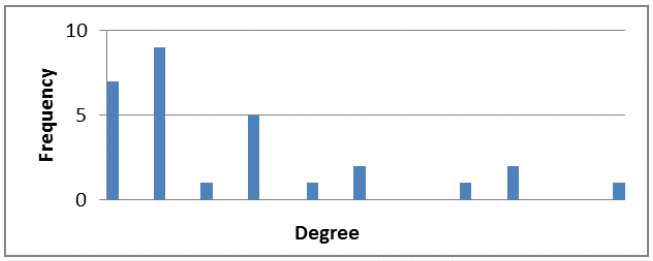

Figure 2. Degree distribution

Table 5 displays the most prominent keywords according to the degree value.

Table 5. Nodes of higher degree and betweenness in the semantic network

Nodes	Degree	Betweenness
Radiotherapy	11	127.484
Chemotherapy	9	-
Biopsy	9	69.930
MRI	8	109.722

*Range of *degree*: 1-11 and Cut-off value: 6;
Range of *betweenness*: 0-127.484 and cut-off value: 55.2

The most prominent keywords are of those treatments and technologies which are most commonly used for cancer.

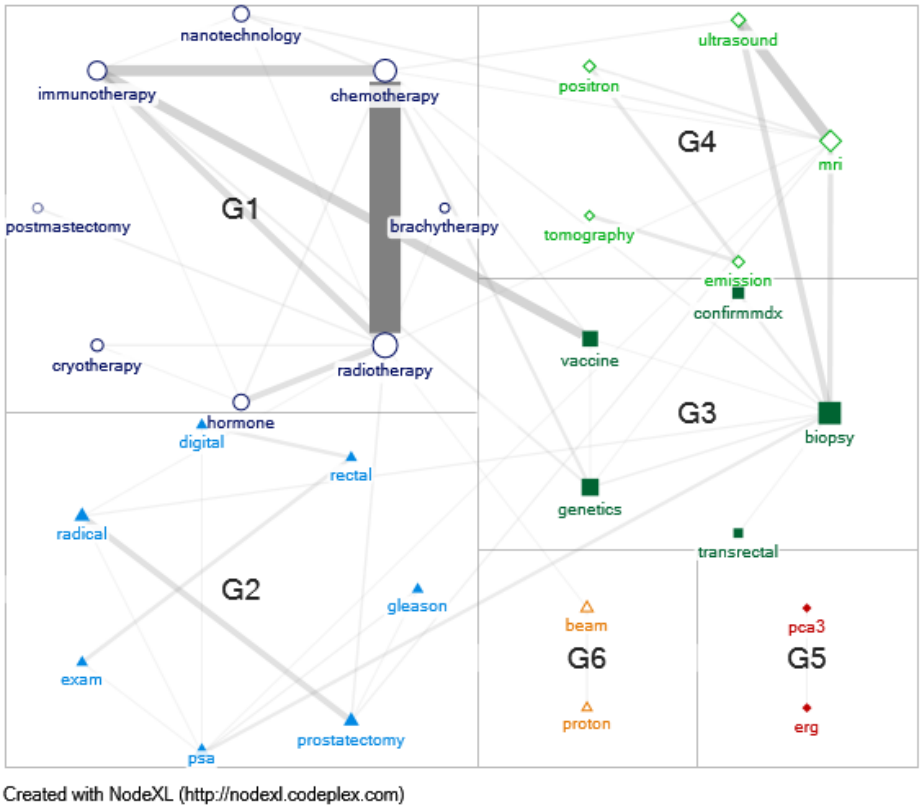

Figure 3. Cluster analysis. Note that the nodes (keyword) of each group are displayed in different colors and shapes. The node size is proportional to the degree, while the edges are thicker and darker for nodes which are paired more often together in tweets, and thinner and lighter for those paired less often

Radiotherapy and *Chemotherapy* are popular choices of treatments for cancer in general, and not only prostate and testicular cancers. While *Biopsies* and *Magnetic Resonance Imaging (MRI)* are commonly used together on the diagnosis phase prior treatment. The fact that the most prominent keywords are of those commonly used technologies and treatments for Prostate and Testicular cancers shows that people tweeting about technology and treatments seem to be aware about the mainstream methods for treating and using technology for Prostate and Testicular cancers.

The betweeness, instead, is a measure of the number of the shortest paths between other nodes that pass through that node. It indicates the node's ability or influence on transfer of items within the network (Freeman, 1979). The higher the betweeness, the most strategic is the keyword because it connects important parts of the graph. Therefore, if the keyword is taken out from the network, all other keywords will be disconnected. This means that keywords with high betweeness are the ones that make the discourse possible. Such keywords symbolize major and central technologies or treatments that people mention to talk about minor and more peripheral technologies or treatments in the discourse.

In our semantic network, the keywords *Radiotherapy, MRI* and *Biopsy* have the highest *betweenness* scores as displayed on table 5. This implies that *Radiotherapy, MRI* and *Biopsy* are strategic 'central' words which are held responsible for the convergence of the network. Due to the presence of these words within the network, people seem to be aware of other minor technologies and treatments' keywords, such as *Cryotherapy, Transrectal Ultrasound* and *ConfirmMDX*. The absence of the keywords *Radiotherapy, MRI* and *Biopsy* would lead to a loss of linkages within the network, and as edges in this case symbolize awareness, this will mean a decrease in awareness of technologies and treatments.

4.2.3 Cluster analysis

Cluster analysis is used to further investigate the semantic network; such analysis serves to identify the sets of keywords which are often used together within tweets. The identification of semantic cluster in the discourse allows to identify the thematic cores of the online discussion on technology and treatments. We run the cluster analysis using the Clauset Newmann algorithgm provided by NodeXL. We identified 6 groups. Figure 3 shows the results.

Starting from the smallest groups, G5 and G6 which consist of only two keywords each, making them dyad clusters in the semantic network. We can see how these small groups refer to very specific topics of discussion. On the one hand, Group 5 consists of the keywords *PCA3* which refers to *Prostate Cancer Antigen 3* and *ERG* which in this case refers

Table 6. Bigger clusters in the semantic network

Group	Number of keywords	Keywords	Theme
G1	8	Radiotherapy Chemotherapy Immunotherapy Cryotherapy Hormone therapy Brachytherapy Postmastectomy Nanotechnology	Types of cancer treatments
G2	7	Digital, Rectal, Exam Gleason Radical, Prostatectomy Prostate-specific Antigen (PSA)	Examination techniques and cancer staging/diagnosis methods
G3	5	Biopsy Genetics Vaccine ComfirmMDX Transrectal Ultrasound (TRUS)	Tests and Diagnosis
G4	5	Positron, Emission, Tomography Magnetic Resonance Imaging (MRI) Ultrasound	Diagnostic Imaging techniques

to the *TMPRSS2:ERG* which are both gene tests that give an indication of the presence of prostate cancer. This explains the pairing of the two keywords in one cluster, as they are both related to genes and are forms of testing for prostate cancer. On the other hand, group 6 contains two keywords which are correlated to proton therapy or also knows as proton beam therapy which is a less common radiation therapy using proton beams instead of x-rays, unlike the conventional and more popular radiotherapy. Proton therapy appears independently probably because it has been introduced to cancer treatments relatively recently (http://www.proton therapy.org/, 2016) in comparison to other therapy methods like chemotherapy and radiotherapy and is not commonly made available in hospitals. This very distinct and small grouping of G5 and G6 shows keywords which are extremely specific, where only few users are able to tweet about such keywords together. The discourse of these two specialized groups can thus be considered a niche discourse as it is very specific.

The analysis of the bigger clusters provides instead evidence of more generic topics of discussion. Table 6 shows the bigger clusters with the corresponding keywords in each cluster, as well as the fundamental theme of each cluster.

G1 is the biggest cluster as it includes 8 keywords, two of which are prominent keywords (Chemotherapy and Radiotherapy). The theme of this group is very broad as it seems to be centered on the types of treatments for cancer, as well as technology types used for different treatments. G4 also contains thematic keywords that portray diagnostic imaging techniques used for the treatment of cancer. G2 and G3 are a little less specific, and they seem to hold keywords related to examination techniques, tests and diagnosis methods.

Consequently, we can conclude that the examination of contents enable us to see that Twitter can be used to portray different themes of technology and treatments of Prostate and Testicular cancers, and to frame the discourse within specific thematic groups. In this case, two themes seem to be dominant: treatment and diagnosis. The cluster analysis show how bigger groups refer to more broad and generic theme of discussion such as treatment (G1), whereas smaller groups are characterized by very specialized topics, such as Genetics tests and Proton Beam Therapy (a specific type of radiation treatment) in G5 and G6. This findings indicates that Twitter users talk more often about generic treatment and show an overall awareness. by contrast, discussion on specific diagnosis appears as a ”niche” discourse that may imply less awareness.

4.3 Content Analysis: Who Tweets about health technology?

Up until this point, we have found that only 0.5% of the UK 2014 *Movember* tweets mention technologies and treatments of prostate and testicular cancers. Within these 0.5%, the prominent keywords used were *Radiotherapy, Chemotherapy, Biopsy* and *MRI*. The cluster analysis of these tweets showed that 6 thematic groups in the semantic network, the largest being very broad and is concerned about cancer treatments while the smallest clusters were more specific.

During the UK 2014 *Movemebr* campaign 1439 users tweeted tweeted about technology and treatment of Prostate and Testicular Cancers related content. This number of users corresponds to 0.4% of the total number of users of this study ($n = 326234$). The profile descriptions of all the 1439 users were manually analyzed and classified using the social iden-

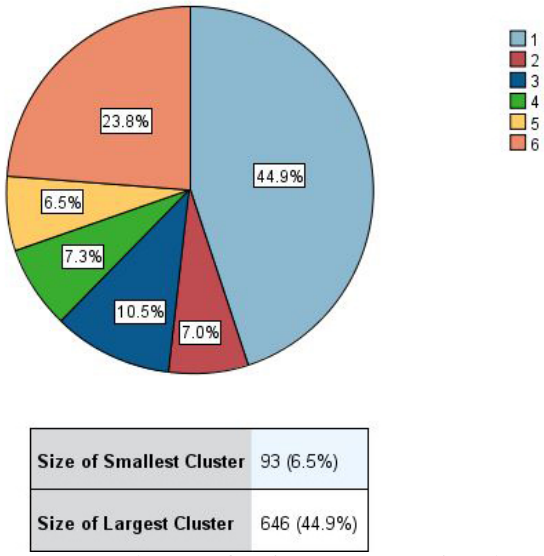

Figure 4. Pie chart showing the six different groups/clusters of Twitter users tweeting about Technology and Treatment of Prostate and Testicular Cancers

Table 7. Group/Cluster with corresponding number of profiles and identity composition

Cluster number	Number of Profiles	Identity
1	646 (368)	Not Applicable (Organizations)
2	101	Occupation and Stigma
3	151	Occupation and Relation
4	105	Occupation and Political
5	93	Occupation and Ethnic/Religion
6	343	Occupation

*Total number of profiles/users was 1439

tity classification system explained in section 2.1. Based on the content analysis analysis made, users tweeting about technology and treatment of Prostate and Testicular Cancers from the UK 2014 *Movember* campaign form six social identity groups or clusters (see Figure 4 and Table 7).

55,1% ($n = 793$) of the profiles belong to social identity categories that are clustered as can be seen on Figure 4 and Table 7. The 'Occupation' identity was predominant (23.8%) when compared to the other four social identity categories. On the other hand, the remaining ($n = 646$) profiles were 'Not Applicable' which represents 44.9% of the total number of profiles. From 646 'Not Applicable' profiles, 57% belonged to Organizations, while the remaining 43% varied in content as some were ambiguous statements and thus belonged to no social identity category.

In summary, it can be mentioned that the occupational identity was predominant. Meanwhile, it can also be stated that a large number of profiles belonging to the 'Relation' identity category tweeted about treatments of family members, friends, and other relations. As for the 'Stigma' category, an observation was noticed which was dominant within the category. Many profiles included terms like 'cancer survivor' which placed the profiles into the 'Stigma' identity category.

In addition to that, many of the Twitter users tweeting about technologies and treatments of prostate and testicular cancers are organizations. By looking at the organizations' tweets contents, it was evident that their purpose of tweeting about technology and treatments in the *Movember* campaign was promoting their products or services which are correlated to health. This finding comes as confirmation of the results obtained from the study about Twitter conversations from the 2013 *Movember* Canada campaign conducted by Bravo & Hoffman-Goetz (2015d). Therefore, organizations made use of the *Movember* campaign online to increase profit and for brand awareness. From this, it can be seen that in many cases Twitter is used as a broadcasting platform for organizations.

To conclude, this social identity results give a clearer idea on the type of users who are involved in discussions of technology and treatments of Prostate and Testicular cancers within the *Movember* campaign online, specifically on Twitter. Twitter users who identify themselves according to their occupations or vocations/interests seem to be more probable to

tweet about technologies and treatments. We find this result interesting because it may imply that people tweeting about technologies and treatments find that the occupation is more important to them when compared to the other identity categories. This assumption can be investigated in further future research.

4.4 Limits and Implications of this research

There are a few limitations in this research. One important limitation to this work is that it only provides results obtained from Twitter conversations of one year (2014) about the *Movember* campaign in the UK. Therefore, it is not possible to generalize the results to a global setting or different times, as results of other countries or years may vary. Similarly, this research analyzed content available only on Twitter and thus the results cannot be extended to all social media platforms.

5. Conclusions

This paper managed to provide answers regarding whether users participating in the 2014 UK *Movember* campaign are tweeting about technology and treatment, and who those users are. Furthermore, the way such users frame their tweets was also investigated.

In the end it can be concluded that there are 1439 users tweeting about technology and treatments. The tweets of such users contribute to only %0.5 of the tweets from the UK 2014 *Movember* campaign, which is a very small percentage indicating that technology and treatments of Prostate and Testicular cancers is not a dominant topic of discussion during the UK 2014 *Movember* campaign on Twitter. In the case where users do tweet about technology and treatment, the fundamental themes mentioned were regarding types of cancer treatments, examination techniques, diagnosis methods and technologies used for diagnostic imaging. Predictably, the most prominent keywords of technology and treatments were of the most commonly used for the diagnosis and treatment of Prostate and Testicular Cancers within the UK. Furthermore, the social identity of individual users tweeting about technology and treatment was predominantly Occupational, as 55.9% of the profiles were classified under the 'Occupation' category of social identity. Moreover, it can be stated that organizations made use of the *Movember* campaign on twitter to promote their products and services, while most individuals used it to express their interest in technology and treatments, or to showcase their topic of research as academics and/or to show support to the campaign and relatives suffering of cancer.

5.1 Recommendations for Future Research

Given the chance to further investigate the presence of Technology and Treatment of Prostate and Testicular Cancers within the conversations of online users, this study would be extended to consider other social media platforms. This would provide an insight of which platform is most suitable to provide awareness or include conversations about technologies and treatments correlated with Prostate and Testicular Cancers. Additionally, further research may be used to evaluate measurable outcome regarding the awareness of such technologies and treatments in the UK, which has not been made for this paper due to time restrictions and lack of resources. This can be investigated by using extra research methods such as surveys and interviews to determine whether social media platforms have an influence on the actual behavior of people, knowledge and awareness of Prostate and Testicular Cancers technologies and treatments.

Additionally, the results obtained concerning the social identity of Twitter users who tweet about technologies and treatments may be implemented to conduct further research. As an example, it may be a possibility that people who view occupation as more important than other social identity categories such as those regarding relations or religion are more probable to talk about technologies online. This hypothesis can kick-start further research.

Acknowledgments

I am highly indebted to the NIKOS research group at the University of Twente, for allowing me access to the Twitter data granted to them in 2014 by Twitter, and therefore having the ability to conduct this research. A Special thanks goes to Dr. Michel Ehrenhard for introducing me to the opportunity to do this study, as well as supervising me and offering guidance.

I would like to also express my special gratitude and thanks towards my direct supervisor Anna Priante, for imparting her knowledge and expertise in this study, and the encouragement provided by her throughout the time of this study.

Last but not least, I would like to thank my family, not only for their initiative in sending me to study abroad, but also for their love and support throughout my bachelor degree years while they are on another side of the planet. Without their moral support none of this would have been possible.

1. Technology and Treatment of Prostate and Testicular Cancers keywords

The following list of keywords was derived after conducting a short review of existing literature related to commonly used and researched technologies and treatments for Prostate and Testicular cancers. In the end, the works of Wasson et al. (1993); Heidenreich et al. (2011) as well as the website of the Cancer Research Society in the United Kingdom (http://www.cancerresearchuk.org/, 2016a,b) has been used to establish the following list of keywords related to Prostate and Testicular cancers technologies and treatments:

- Gleason
- Prostate cancer antigen 3 (PCA3)
- Radiation therapy / Radiotherapy
- Chemotherapy
- Hormone therapy
- Conformal radiation therapy (CRT)
- Intensity modulated radiation therapy (IMRT)
- Proton beam radiation
- Nanotechnology
- Radiofrequency
- Ultrasound
- Transrectal ultrasound(TRUS)
- Color Doppler ultrasound
- Biopsy
- MRI / scanner / mri-scanner
- Positron-emission tomography (PET)
- Multiparametric MRI
- Brachytherapy
- Radical prostatectomy
- Genetics/genes
- Immunotherapy
- Hormonal medicine ”5-alpha reductase inhibitors”
- Prostate-specific antigen (PSA)
- 4Kscore
- Digital rectal exam (DRE)
- TMPRSS2:ERG
- ConfirmMDx
- Vaccine treatment
- Bone directed treatmemt
- Cryotherapy

References

Ashforth, B. E., Harrison, S. H., & Corley, K. G. (2008). Identification in organizations: An examination of four fundamental questions. *Journal of Management*, *34*.

Ashforth, B. E., Schinoff, B. S., & Rogers, K. M. (2016). Identification in organizations: An examination of four fundamental questions. *Journal of Management Review*, *41*.

Bravo, C. A., & Hoffman-Goetz, L. (2015a). Social media and men's health: A content analysis of twitter conversations during the 2013 movember campaigns in the united states, canada, and the united kingdom. *American Journal of Men's Health*, *1557988315617826*, 1.

Bravo, C. A., & Hoffman-Goetz, L. (2015b). Tweeting about prostate and testicular cancers: Do twitter conversations and the 2013 movember canada campaign objectives align? *Journal of Cancer Education*.

Bravo, C. A., & Hoffman-Goetz, L. (2015c). Tweeting about prostate and testicular cancers: Do twitter conversations and the 2013 movember canada campaign objectives align? *Journal of Cancer Education*.

Bravo, C. A., & Hoffman-Goetz, L. (2015d). Tweeting about prostate and testicular cancers: What are individuals saying in their discussions about the 2013 movember canada campaign? *Journal of Cancer Education*.

Deaux, K., Reid, A., Mizrahi, K., & Ethier, K. A. (1995). Parameters of social identity. *Journal of Personality and Social Psychology*, *68*.

Drieger, P. (2013). Semantic network analysis as a method for visual text analytics. *Procedia-social and behavioral sciences*, *79*, 4.

Freeman, L. (1979). Centrality in social networks: Ii. *Social networks*, *2*.

Hansen, D., Shneiderman, B., & Smith, M. (2009). Analyzing social media networks: Learning by doing with nodexl. *Network Analysis*.

Heidenreich, A., Bellmunt, J., Bolla, M., Joniau, S., Mason, M., ... Zattoni, F. (2011). Eau guidelines on prostate cancer. part 1: screening, diagnosis, and treatment of clinically localised disease. *European urology*, *59*.

http://www.cancer.org. (2016a). *What's new in prostate cancer research?* (Retrieved from *hhttp : //www.cancer.org/cancer/prostatecancer/detailedguide/prostate − cancer − new − research*)

http://www.cancer.org. (2016b). *What's new in testicular cancer research and treatment?* (Retrieved from *http : //www.cancer.org/cancer/testicularcancer/detailedguide/testicular − cancer − new − research*)

http://www.cancerresearchuk.org/. (2016a). *Prostate cancer.* (http://www.cancerresearchuk.org/about-cancer/type/prostate-cancer/)

http://www.cancerresearchuk.org/. (2016b). *Testicular cancer.* (http://www.cancerresearchuk.org/about-cancer/type/testicular-cancer/)

http://www.proton therapy.org/. (2016). *Background: The voice of the proton community.* (http://www.proton-therapy.org/backgrou.htm)

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*.

Jacobson, J., & Mascaro, C. (2016). Movember: Twitter conversations of a hairy social movement. *Social Media+ Society*, *2*.

Krikorian, R. (2014). *Twitter #datagrants selections.* (Retrieved from *https : //blog.twitter.com/2014/twitter − datagrants − selections*)

Krippendorff, K. (2004). Reliability in content analysis. *Human Communication Research*, *30*, 411-433.

Lang, T. (2011). Advancing global health research through digital technology and sharing data. *Science*, *331*, 714-717.

Lee, Y., Lee, J., & Lee, Z. (2006). Social influence on technology acceptance behavior: self-identity theory perspective. *ACM SIGMIS Database Press*, *37*.

Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer: analyzing and comparing opinions on the web. *Proceedings of the 14th international conference on World Wide Web*.

Movember Foundation. (2014). *Movember 2014 global annual report.* (`https://cdn.movember.com/`)

Pavan, E. (2012). Frames and connections in the governance of global communications: A network study of the internet governance forum. *Lexington Books*.

Percheski, C., & Hargittai, E. (2011). Health information-seeking in the digital age. *Journal of American College Health*, *59*.

Pew Research Center. (2014). *Social media update 2014.* (Retrieved from *http : //www.pewinternet.org/files/2015/01/PIsocialMediaUpdate20144.p*

Prasetyo, N. D., Hauff, C., Nguyen, D., Hiemstra, D., & van den Broek, T. (2015). What makes or breaks a health fundraising campaign on twitter? *6th International Workshop on Health Text Mining and Information Analysis Workshop*.

Priante, A., Hiemstra, D., van der Broek, T., Ehrenhard, M., & Need, A. (2016). Classifying social identities on twitter. *Work-in-progress manuscript*.

Ratzan, S. (2011). Our new "social" communication age in health. *Journal of health communication*, *16*.

Rosemary Thackeray, C. G.-C. S. R., Scott H Burton, & Draper, C. R. (2013). Using twitter for breast cancer prevention: an analysis of breast cancer awareness month. *BMC Cancer*, *13*.

Scott, C. (2007). Communication and social identity theory: Existing and potential connections in organizational identification research. *Communication Studies*, *58*.

Stryker, S., & Burke, P. J. (2000). The past, present and future of an identity theory. *Social Psychology Quarterly*, *63*.

Tait, D. (2015). *To improve radiotherapy for patients, we need to raise its profile.* (http://scienceblog.cancerresearchuk.org/2015/08/28/to-improve-radiotherapy-for-patients-we-need-to-improve-its-profile/)

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. the social psychology of intergroup relations. *Annual review of psychology*, *33*.

Tamburrini, N., Cinnirella, M., Jansen, V. A., & Bryden, J. (2015). Twitter users change word usage according to conversation-partner social identity? *Social Networks*, *40*, 84.

UTwente. (2014). *Twitter data grant cancer early detection campaigns on twitter.* (https://www.utwente.nl/en/news/!/2014/4/329288/twitter-data-grant)

van der Veen, H., Hiemstra, D., van den Broek, T., Ehrenhard, M., & Need, A. (2015). From slacktivism to activism: participatory culture in the age of social media. *arXiv preprint arXiv:1508.02483*.

Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications. *Cambridge: Cambridge University Press*.

Wasson, J. H., Cushman, C. C., Bruskewitz, R. C., Littenberg, B., Albert G. Mulley, J., Wennberg, M. J. E., ... the Prostate Disease Patient Outcome Research Team (1993). A structured literature review of treatment for localized prostate cancer. *Archives of Family Medicine*, *2*.

Younis, E. (2015). Sentiment analysis and text mining for social media microblogs using open source tools: An empirical study. *International Journal of Computer Applications*, *112*.