Big Social Data Analytics for Public Health: Facebook Engagement and Performance

Nadiya Straton[1], Kjeld Hansen[1,2,3], Raghava Rao Mukkamala[1], Abid Hussain[1], Tor-Morten Grønli[2]
Henning Langberg [3] and Ravi Vatrapu[1,2]
{nsr.itm,rrm.itm,ah.itm,rv.itm}@cbs.dk[1], {hankje,tmg}@westerdals.no[2], henninglangberg@gmail.com[3]
[1]Centre for Business Data Analytics, Dept. of IT Management, Copenhagen Business School, Denmark
[2] Mobile Technology Laboratory, Westerdals Oslo School of Arts, Communication and Technology, Norway
[3]CopenRehab, Dept. of Public Health, University of Copenhagen, Denmark

***Abstract*—In recent years, social media has offered new opportunities for interaction and distribution of public health information within and across organisations. In this paper, we analysed data from Facebook walls of 153 public organisations using unsupervised machine learning techniques to understand the characteristics of user engagement and post performance. Our analysis indicates an increasing trend of user engagement on public health posts during recent years. Based on the clustering results, our analysis shows that Photo and Link type posts are most favourable for high and medium user engagement respectively.**

I. Introduction

Recent research shows growing number of studies and articles about social media and health care. Growing number of healthcare organisations and individuals are realising the benefits of using social media to communicate health information. Their main purposes include to train medical personnel, provide information to patients and allow rapid communication in times of crises, through social media sites [1]. Social media provides new opportunities for interaction and distribution of information within and across organisations, which results in new kinds of socially mediated organisations [2]. As such, knowledge on how organisations spread - and how users interact with health information through social media and mobile computing will become increasingly important in the near future.

This emphasis on understanding the socio-technical interactional contexts of users and health care organisations is already evident in the new public health paradigm in general and the field of health informatics in particular. Public health has traditionally been understood through its unit of analysis i.e. the public; However, the new public health paradigm *goes beyond an understanding of human biology and recognizes the importance of those social aspects of health problems, which are caused by life styles. In the new public health the environment is social and psychological as well as physical* [3] as quoted in [4]. Within the public health paradigm, the field of health informatics deals with *the structures and processes, as well as the outcomes involved in the use of information and information and communication technologies (ICTs) within health* [5, p.501-502]. Situated within new public health and health informatics, this paper investigates the distribution and user engagement with health information on the Facebook walls of the official portals of Public Healthcare Services.

Social and health scientists have shown considerable interest in investigating the importance of the connection between our social lives and health situations. Christakis and Fowler [6] have studied how our social networks can influence our health situation as a consequence of *everything we think, feel, do, or say can spread far beyond the people we know...they can help us to achieve what we could not achieve on our own* [6, p.30-31]. Facebook is amongst the leading social media network channels globally and has approximately 1.5 billion active users on monthly basis. Even though users tend to be active on more than one social media network channels, most people consider Facebook their social media home. The widespread societal and individual adoption of Facebook has led to a new kind of relationship between people and information [7]. We term the health information shared on social media platforms as *Socially Shared Health Information* (SSHI) and seek to understand the organisational rationale for using SSHI and user interactions with SSHI with special focus on analysing the official Facebook wall of Public Health Organisations. In this research work, we explore the following research question and propositions:

> *What are the key characteristics of post performance of the Socially Shared Information on the Facebook walls of Public Health organisations?*

The following are the propositions for the research question.

- Healthcare Facebook walls contain interesting information for public which can be indicated by increase of user interactions such as shares, likes and comments.
- Out of several attributes that indicate engagement of Facebook posts, *Page Likes* is likely to be related to the number of *Post Likes* and *Post Shares*.
- Facebook posts with visual content such as *photo* and *video* are supposed to be more engaging than posts with only text content.

The rest of the paper is organised as follows. In the next section, we present related work and then a brief description of the research methodology adopted for this work is presented in sec. III. The main results of our analysis are presented in sec. IV and finally we conclude in sec. V.

II. Related Work

First of all, Hamm et al. [1] defines several groups of social media cites by their content and defines social networking site Facebook as one of the most actively used in the USA. However they noticed that in spite of wide spread usage of social media among the users, adoption of social media tools by the health care professionals around the globe is relatively slow. Additional findings were concluded that main use of social media by health care professionals was aimed at achieving communication and knowledge rather than skill objective. YouTube has been used for promotion of information about cancer screening as well as obesity, Twitter in the design of interventions about prenatal health promotion, while Facebook was used in interventions related to sexual health issues [8]. There is also considerable number of standalone health focused social networking applications used for various chronic diseases like diabetes [8]. Hamm et al. [1] argues that preference for more tailored applications over Facebook is understandable if one takes into account issues related to professionalism. Authors highlight further that in spite of benefits Facebook brings for the health care professionals, it also carries potential risks. Privacy of individuals who carry a disease and turn to Facebook for information might be jeopardized, as this data is gathered without their consent. However there are low cost and a benefit to transcend geographical boundaries while providing information through social media [9]. There have always been debates for and against use of social media in health care; therefore issue of professional boundaries and unbiased sharing on Facebook is valid in many studies.

Moreover, with the increasing amount of health care information available through social media there is a great possibility for new insights by using bigdata analytics. In this paper we apply machine learning techniques on big social data of public health Facebook walls to gain insights into user engagement and post performance. Facebook is traditionally considered a media site for friends, where information sharing and usage might be related and biased to the density of the circle of friends rather than a place to give independent medical advice and spread disease specific information.

Study on *Supportive responses to social sharing on Facebook* shows that people with larger and diverse friend networks with larger potential audiences behave differently on Facebook when they share status updates and emotion icon. Therefore the higher and more diverse the audience of friends on Facebook is, the higher the likelihood of positive status updates. Diversity and large audience means that people tend to limit themselves to more diplomatic topics and happy news [10].

However possibly limited amount of studies of health care and Big data on Facebook could be due to the different types of content from video, picture, text to discrete numbers that characterize different features. These intermix might contribute to the difficulty of information gathering and analytics. This research focuses only on studying the impact of discrete type features from health care posts on Facebook and the efficiency of communicating and spreading health care information. Our research methodology is using statistical methods to make first introduction into the data and then application of the K-Means algorithm to cluster or group health care data on Facebook into popularity buckets. Relevant features for each of the clusters are studied further.

III. Research Methodology

Start date: **2006-01-01** End date: **2015-12-30**		
Number of Facebook Walls: **153**		
Activity	**No. of Actions**	**Unique Actors**
Facebook Page Likes	10, 476, 523	–
Facebook Posts	280, 534	101, 351
Post Shares	4, 225, 739	–
Likes on Posts	24, 331, 261	7, 129, 957
Comments	1, 734, 154	788, 297
Likes on Comments	1, 507, 687	493, 266
Comment Replies	208, 512	100, 379
Likes on Comment Replies	176, 920	88, 202
Total	42, 941, 330	7, 531, 865[1]

Table I
Overall Statistics of Public Health Facebook Dataset

In this section, first we describe the dataset and data collection method, then we will introduce the statistical methods and data clustering method in the subsequent sub sections.

A. Dataset Description

Category	Remarks
Blogger	Influential health blogs operated by individuals
Campaign	Time-limited health campaigns (one-time/recurring)
Community	Voluntary union of individuals join together for one or more health agendas
Disease specific event	Disease related events/health care situation that is partially or fully arranged through Facebook
Disease specific organization	Organisations that are focused on one or more health/sickness situations
Foundation	Economically independent entities with clear objective - with specific health/sickness focus.
Governmental	Public-faced governmental entities with a specific health/sickness focus
Individual	Private individuals, known and respected for their specific health related expertize. (and prefer to communicate through FB - rather than a blog.)
Media	Media outlets in general - with or without a specific health/sickness focus.
NGO	Non-governmental organisations with a general health/sickness focus.
Pharma education	An educational program with a public health focus
Pharma	Pharmaceutical company, which produces products related to common health/sickness situations
Research	Public faced research projects, programs or initiatives with a health/sickness focus
Team	Sports team with a specific health/sickness focus
TV	TV channels and/or programs in general or with a specific health/sickness focus

Table II
Category Type Descriptions

Data from 153 public Facebook walls belonging to various public health organisations is collected using the Social Data Analytics Tool (SODATO) [11]. These walls include

[1]Total unique actors for the whole dataset

national as well as international agencies, organisations as well as individual bloggers as categorised in table II. The total dataset contains information about around 43 million Facebook actions that happened during a time period of 10 years as shown in table I. Majority of actions are *Likes on Posts* (around 55%) and the dataset contains around 280 000 Facebook posts. Around 34% of dataset are *Facebook Page Likes* and *Post Shares* and Facebook does not provide user information in respect of these items. In the entire dataset, there are around 7.5 million unique users and as one can notice that from table I that the prominent action performed by the users is *like* action on posts.

B. Descriptive Statistics

The primary goal is apply unsupervised data mining/machine learning techniques (such as data clustering) on the Facebook data to understand the characteristics/attributes of posts that lead to high user engagement and performance. The dataset initially included sixteen direct/indirect attributes and additionally nine derived attributes (such as day of the week, month, year, time of the day, season, is a holiday, post engagement in hours) were generated and added to the data.

All attributes are discrete and there are no missing values. Basic summary statistics of the attributes shows how dispersed are the attributes in relation to the mean, median, min and max values as shown in figure 1.

Attributes	Mean	Std.	Min	Max	Median	75%
Page Likes			22	1608611	110106	248250
Talking About			0	35857	4027	6841
Post Share	15	246	0	66977	0	2
Post Likes	85	756	0	141419	3	23
Comments	6	68	0	18576	1	3
Comment Likes	5	60	0	9774	0	1
Comment Reply	1	10	0	4120	0	0
Comment Reply Like	1	16	0	5639	0	0

Figure 1. General Statistics of Total dataset

There are some issues with data outliers, as attributes are dispersed in general and some of the attribute values are really too high in relation to the majority of the data set, therefore three such outlier values (120,083, 141,419, 130,579) from *Post Likes* were excluded from analysis. As seen from figure 1, for the majority of attributes, the mean value is not equal to the median, which suggests that the data is skewed, which means that the values are not symmetrically distributed around the mean, and therefore not normally distributed.

1) Correlation between attributes: To define the measure of popularity all measures of post engagement were taken into account as shown in figure 2: Page Likes, Talking About, Post Share, Post Like, Comments, Comment Likes, Comment Replies, Comment Reply Likes. The most direct method to see how popular certain health care post and what feedback is gets is to look into Post Like, Post Share and Comment. However to find if these attributes are related at all or if they are mostly independent measures, correlation statistic between attributes was applied. Moreover to define one concept of popularity/post engagement with several attributes it is important to see those that affect direct methods of post engagement the most.

Initial proposition assumed that there must be relation between engagement of health care wall through number of *Page Likes* and active *Talking About* users and actual Posts. Number of users signed up to the page might be involved through Comments, Post share and Post like activities and the higher the number of users signed up to health care wall the higher the activity around the Post. However results showed different picture, were walls with very high number of users used to receive very low engagement per Post. These finding can be evidenced and further explained by low correlation between *Page Like*, *Talking About* and *Post Like*, *Post Share* attribute. Correlation between Facebook Wall engagement attribute and Post engagement attribute is 0.05. In its turn *Page Like* and *Talking About* have correlation attribute of 0,79, which is very close to 1, perfect correlation. *Post Share* is the best correlated with itself and *Post Like* (0.24). Post Like has highest correlation with *Comment Like* (0.30), *Comment* (0.23) and *Post Share* (0.25) and least correlated with Comment Replies and Comment Reply Likes.

Attributes	Page Likes	Talking About	Post Share	Post Like	Comments	Comment Like	Comment Reply	Comment Reply Like
Page Likes	1	0.79	0.02	0.05	0.00	0.00	0.00	0.00
Talking About	0.79	1	0.03	0.06	0.03	0.03	0.03	0.02
Post Share	0.02	0.03	1	0.25	0.18	0.24	0.20	0.14
Post Like	0.05	0.06	0.25	1	0.23	0.30	0.16	0.13
Comments	0.00	0.03	0.18	0.23	1	0.44	0.50	0.25
Comment Likes	0.00	0.03	0.24	0.30	0.44	1	0.58	0.66
Comment Replies	0.00	0.03	0.20	0.16	0.50	0.58	1	0.61
Comment Reply Likes	0.00	0.02	0.14	0.13	0.25	0.66	0.61	1

Figure 2. 2 Correlations between Attributes

With uncorrelated data it might be very difficult to perform unsupervised learning method of clustering, therefore only four attributes most correlated attributes were chosen to represent Post popularity/engagement: *'Post Share'*, *'Post Like'*, *'Comment'* and *'Comment Like'*. The proposition for the relation between Facebook health care *Page Like* and *Post Like* is not supported from this finding.

C. Data Clustering

Clustering [12] is an *Unsupervised* data mining technique to identify groups of similar data items, primarily to discover patterns and interesting correlations that exist in large datasets. There are many clustering algorithms [13] to achieve partitioning of data into a predefined number of clusters with the help of an objective function to compute membership values of each data item, usually based on similarity measures such as distance. One of such popular algorithms is *K-Means* algorithm [14], which has been used extensively in many scientific and industrial applications [13]. *K-Means* algorithm uses a objective function that partitions the dataset in such a way that the sum of the squared errors between centroid (or empirical mean) of a cluster partition to its data points is minimised.

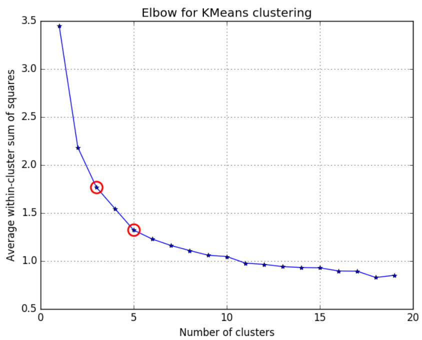

Figure 3. Number of clusters for Facebook dataset

1) Number of Clusters: : Finding the optimal number of clusters for a given dataset is a challenging problem in data clustering [15]. There are several automatic methods for determining the number of clusters, one such measure is based on average error within-cluster dispersion. The error measure decreases monotonically with the increase of number of clusters and for some cluster number k onwards the decrease in average within-clustering dispersion flattens noticeably (also known as Elbow). The number of clusters verses average within-clustering dispersion for our Facebook dataset is shown in figure. 3 and for our analysis, we have chosen the number of clusters as 3.

IV. Results

A. Analysis of Total Dataset

When comparing clusters that were assigned with K-Means algorithm and classes with defined non-overlapping boundaries than we have a clear picture of division for popular and less popular posts. Almost 30% of the High engagement posts are higher or equal to the 727 engagement/popularity value comprised of *'Post Like'*, *'Post Share'*, *'Comment'* and *'Comment Like'*. Cluster with medium level of popularity contains 48.75% of the posts, which lay within 23 - 88 value boundary. Moreover 66.65 % of the posts in the Low valued cluster belong to the 0 - 4 post value boundary and represent 41.7% of the total data, 32.52% of the posts belong to the 4 - 23 value boundary and less than one percent belong to the 23-88 value boundary. Therefore, for majority of the posts, sum of *'Post Like'*, *'Post Share'*, *'Comment'*, *'Comment Like'* values is ≤ 4.

Classes with Total Value Boundaries								
Values	**Clusters**	**4 -- 23**	**88 -- 374**	**0 -- 4**	**374 -- 727**	**23 -- 88**	**727 = <**	**Grand Total**
0 - 404	Low	32,52%	0,01%	66,65%		0,81%		100,00%
9 - 6546	Med	33,62%	16,99%		0,52%	48,75%	0,11%	100,00%
52 - 73948	High		47,16%		21,55%	2,71%	28,58%	100,00%
	Grand Total	**29,50%**	**9,44%**	**41,76%**	**2,35%**	**14,01%**	**2,95%**	**100,00%**

Figure 4. Classes in comparison to clusters

Having a better understanding of grouping of post attribute values into 3 different clusters and what they represent, leads to better understanding of the possible differences between the features that belong to each of those clusters. When mining for the features of the total data some of the features are represented evenly and some show distinctly higher values for one cluster and not the other. Table 5 presents feature values for high, medium and low clusters.

	Engagement Cluster		
	High engagement posts	Medium engagement posts	Low engagement posts
Class Boundaries	**88 -- 374**	**23 -- 88**	**0 -- 4**
Year	2015	2015	2011
Month	Oct	Oct	Oct
Day of Week	Tue	Tue	Wed
Time of Day	10-16	10-16	16-22
Season	Autumn	Autumn	Autumn
Country	Denmark	Norway	USA
Level	National	National	National
Category Type	Disease Specific org.	Disease Specific org.	Disease Specific org.
Holiday/Not	Not a Holiday	Not a Holiday	Not a Holiday
Post Type	Photo	Link	Status
Avg. Hour Span	502	125	37

Figure 5. Feature representation of Overall Dataset

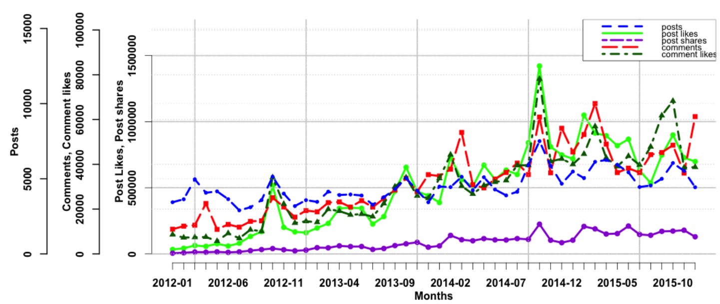

Figure 6. Temporal Distribution of Posts, Shares, Post Likes, Comments and Comment Likes of Total dataset

The increase of post engagement over the temporal dimension is shown in figure 6, which supports propositions, where health care organisations managed to capitalize on social media spread over years. In Figure 5, biggest share of high performance posts are posted recently in 2015 (the latest year), as well as medium performance posts. Whereas the highest share of low engagement posts was seen earlier in 2011. It can be explained also by the fact that overall engagement on Facebook with the 153 walls / health care organization represented in our data set was rather low prior to 2013.

Moreover, the highest percentages of posts around 30% are posted in autumn, with the highest percentage of almost 12% posted around October across all 3 clusters. Here all three clusters show similar trend and it suggests that season or month is not decisive in how the visible post can be and therefore not an influential feature for higher engagement. However this information still gives interesting insight into activity of health care organisations on Facebook.

A total 98% of health care posts are posted during Non-Holiday season (*Holiday* season includes Christmas, New Year and Easter holiday, which are days off traditionally in Europe and USA). The share of data between organisations

		Class Boundaries		Year		Post Type		Avg. Hour Span	
		Cluster							
Engagement	Total Data Set Share	High	Low	High	Low	High	Low	High	Low
Blogger	**0,6%**	**88 -- 374**	**0 -- 4**	2015	2014	Link	Status	104	53
Campaign	**2,1%**	**727 -- =<**	**0 -- 4**	2015	2014	Photo	Status	394	48
Community	**0,3%**	**88 -- 374**	**0 -- 4**	2015	2014	Photo	Link	252	9
DS Event	**10,5%**	**88 -- 374**	**0 -- 4**	2015	2015	Photo	Status	574	22
DS Organization	**67,6%**	**88 -- 374**	**0 -- 4**	2015	2011	Photo	Status	514	35
Foundation	**0,1%**	**88 -- 374**	**0 -- 4**	2015	2013, 11	Photo	Status	47	10
Governmental	**6,6%**	**88 -- 374**	**0 -- 4**	2015	2014	Photo	Link	585	88
Individual	**1,4%**	**88 -- 374**	**0 -- 4**	2013	2015	Photo	Status	472	51
Media	**3,8%**	**88 -- 374**	**0 -- 4**	2015	2012	Link	Link	149	27
NGO	**1,1%**	**88 -- 374**	**0 -- 4**	2013	2015	Photo	Link	220	82
PH degree	**0,1%**	**88 -- 374**	**0 -- 4**	2015	2012	Link	Link	3	2
Pharma	**2,7%**	**4 -- 23**	**0 -- 4**	2015	2013	Photo	Status	581	114
Research	**1,8%**	**23 -- 88**	**0 -- 4**	2013	2012	Photo	Link	246	38
Team	**0,8%**	**727 -- =<**	**0 -- 4**	2014	2013	Photo	Status	156	42
TV	**0,4%**	**88 -- 374**	**0 -- 4**	2014	2013	Photo	Status	395	187
Total Data Set Values	**100,0%**	**88 -- 374**	**0 -- 4**	2015	2011	Photo	Status	502	37

Figure 7. Category type feature representation

and individuals in different countries is comprised of: 18% of data from Denmark, 17% of data from Norway, 20,5% from United Kingdom and 31% from the USA.

Moreover, 64% of health care posts represent *'Not for Profit Organization'* as wall specific category defined by Facebook and is representative in all three clusters. Similarly, 68% represent Disease Specific Organization and refers to organisations, which are focused on one or more health specific issues. Highest share of *'Disease Specific Organization'* is seen across all three clusters. However there are more distinct features for high engagement posts in comparison to low engagement posts, in *Post Type*, *Hour Span*, *Time of the Day*, *Day of the Week* and *Year* features. The highest share of high and medium engagement post are posted on Tuesday and highest share of low engagement posts are posted on Wednesday.

With respect post type, high post engagement is almost in 50% of the cases is *photo* message, while 35% of the high engagement posts are represented with the Link type message. In low engagement posts only in 8% of the cases by the Photo and in 65% of the cases by the Status. Medium engagement cluster post follow similar trend, where photo type posts comprise only in 37% of the posts. Therefore, *Photo* attracts more attention for health care posts on Facebook than *Link* since with its share decrease, decreases post engagement. *Average Hour Span*, which represents a difference between *Post Create* and *Post Update* date, also shows increase, with increase in Post Engagement. *Average Hour Span* for high engagement posts is a little over 500 hours, which is around 20 days. With increase in Facebook post presence the visibility of the post naturally increases and supports the propositions.

Day of the week also shows that high engagement posts are mainly posted on Tuesday and low engagement Posts are mostly posted on Wednesday. In general low activity around posts falls on the weekend, as 67.64% of the posts are *Disease Specific Organization* and less than 1% are posts by the individual users. Individuals might not be limited to working schedule as organisations are limited to working week. Around 44% of high engagement and medium engagement posts are published at the time slot between 10.00 and 16.00, which gives a chance for individual users to comment right after working hours or during lunch. Only 9% of high performance posts are posted during 22.00 - 04.00, where as 20.4% of low engagement posts are posted during nighttime. Time difference around the globe is the reason why organisations are posting during nighttime CET and partially why the majority of low engagement posts are posts from the USA. In general, Norwegian and Danish organisations and individual posts represent almost 40% of the post volume and also they show a very good post engagement.

B. Analysis based on Category Types

Category Types follow similar trend as total data for some features. Since Disease Specific Organization represents the majority of our data (67%), its features are the most representative in the total data set. Therefore to understand feature for each category better, we applied K-Means algorithm to all the categories individually. Results in figure 7 shows number of features represented in high and low clusters which also show similar trend as that of features for the total dataset.

From the figures 5&7, we could notice that boundaries for high and low clusters is more or less the same for total dataset and category-wise analysis. Similarly, highest share of high engagement posts across all the fifteen categories is seen in 2015, whereas highest share of the low engagement posts is seen in previous years. Even though there is a variance in years across all the categories, the trend is clear, there was relatively higher share of low engagement posts in previous years. The reason for that could be relatively lower attention to health care related Facebook posts in previous years.

Post activity pattern from 2006 to 2015 shows that only 1.5 percent of the posts were posted from 2006 to 2009 and majority 77 percent were posted during last four years from 2012 to 2015 of the data set. Even though there was a clear increase in health care post activity from 1.3 per cent in 2009 to 15.6 per cent in 2011, there was lack in real response from the people using social media.

Frequency analysis and K-Means clustering results show that posts from earlier years are mainly represented by low engagement cluster and therefore those posts had a very low user attention/engagement. This trend is represented and seen through years from 2006 to 2012. From 2013 and onwards there is clearly higher attention to the posts from Facebook users, which is represented by increase of share in high and medium clusters. Better post engagement can be explained by increase in average hour span, due to the fact that post is seen online more frequently. Another reason could be a post with similar message from previous years is likely to get more attention in 2015 than in years prior to 2013. Therefore, the higher the average span of engagement the better the chance for the post to gain more attention due to the increase in popularity of health care posts on Facebook in recent years.

Our analysis shows that 'Photo' type posts and 'Link' type posts are most favourable for high and medium engagement respectively. In any case only small fraction of Photo type posts are part of the low engagement post, which suggests importance of communicating the message through the visual means in order to attract more attention. Moreover, none of the medium engagement posts are represented by Status posts and mostly contain Link or Picture. Very often through the Link message one receives a visual representation of the shared content in the form of a picture.

There is a high share of high engagement and medium engagement posts that are published in a time span from 10-16, whereas low engagement posts are published later in the evening across all the categories. However there are some categories, which show the opposite trend. It might be hard to conclude with certainty that health care posts posted from 10-16 are necessary high engagement posts, without taking into account what category type they belong to. In 8 out of 15 categories (table II) such as: Disease Specific Event, Blogger, Community, Foundation, Media, NGO, Pharmaceutical Degree and Team all posts are published in the same time slot. Therefore time period of the post will not contribute to the conclusive results when analyzing post popularity. This is contrary to the previous finding from the total data analytics where major share, almost 50% of medium and 50% of high performance posts were published in the time slot from 10-16 and low performance in the evening time slot. Furthermore month, day of the week, season and country are not decisive features in placing post into certain cluster in general.

V. Conclusion and Future Work

In this paper, we have analysed data pertaining to 153 public Facebook walls that belongs to various public health organisations using data Clustering and other statistical methods. Our analysis shows that, to achieve better engagement, organizations must represent their posts through Photo or Link rather than other types of information sharing techniques. Therefore, to be more efficient in communicating health related issues and create awareness among users, the organisations must avoid status updates and think about better visual representations. Status updates that are posted by a famous persons actively followed by a large number of Facebook users.

Posts that are visible and active for some time attract more user engagement. Also we noticed a trend of increased user attention during the recent years towards public health Facebook walls. As Facebook is widely adapted by the users and therefore, health care organisations should actively and effective communicate their messages through social media to reach wider audience. A clear indication of increase in health care posts engagement on Facebook during 2014 and 2015 would indicate the same. As part of our future work, we will analyze of the textual content of the public health posts by applying machine learning and supervised classification techniques for domain-specific models from public health such disease specific models, emotions and so on.

References

[1] M. P. Hamm, A. Chisholm, J. Shulhan, A. Milne, S. D. Scott, T. P. Klassen, and L. Hartling, "Social media use by health care professionals and trainees: a scoping review," *Academic Medicine*, vol. 88, no. 9, pp. 1376–1383, 2013. 1, 2

[2] R. Vatrapu, "Understanding social business." in *Emerging Dimensions of Technology Management*. Springer, 2013, pp. 147–158. 1

[3] J. Ashton and H. Seymour, *The New Public Health: The Liverpool Experience*. Buckingham: Open University Press, 1992. 1

[4] A. Petersen and D. Lupton, *The new public health: Health and self in the age of risk.* Sage Publications, Inc, 1996. 1

[5] P. A. Bath, "Health informatics: current issues and challenges," *Journal of Information Science*, 2008. 1

[6] N. A. Christakis and J. H. Fowler, *Connected: The surprising power of our social networks and how they shape our lives*. Little, Brown, 2009. 1

[7] K. Kunst and R. Vatrapu, "Towards a theory of socially shared consumption: Literature review, taxonomy and research agenda," in *Proceedings of the European Conference on Information Systems (ECIS), Tel Aviv, Israel*, 2014. 1

[8] P. Balatsoukas, C. M. Kennedy, I. Buchan, J. Powell, and J. Ainsworth, "The role of social network technologies in online health promotion: A narrative review of theoretical and empirical factors influencing intervention effectiveness," *Journal of medical Internet research*, vol. 17, no. 6, 2015. 2

[9] C. Hawn, "Take two aspirin and tweet me in the morning: how twitter, facebook, and other social media are reshaping health care," *Health affairs*, vol. 28, no. 2, pp. 361–368, 2009. 2

[10] M. Burke and M. Develin, "Once more with feeling: Supportive responses to social sharing on facebook," in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 2016, pp. 1462–1474. 2

[11] A. Hussain, R. Vatrapu, D. Hardt, and Z. Jaffari, "Social data analytics tool: A demonstrative case study of methodology and software." in *Analysing Social Media Data and Web Networks*, M. C. Rachel Gibson and S. Ward, Eds. Palgrave Macmillan, 2014 (in press). 2

[12] A. K. Jain, "Data clustering: 50 years beyond k-means," *Pattern recognition letters*, vol. 31, no. 8, pp. 651–666, 2010. 3

[13] P. Berkhin, "A survey of clustering data mining techniques," in *Grouping multidimensional data*. Springer, 2006, pp. 25–71. 3

[14] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979. 3

[15] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 2, pp. 411–423, 2001. 4