

Deep Analyzing Public Conversations: Insights from Twitter Analytics for Policy Makers

Nimish Joseph, Purva Grover, Polaki Rao, P. Ilavarasan

► **To cite this version:**

Nimish Joseph, Purva Grover, Polaki Rao, P. Ilavarasan. Deep Analyzing Public Conversations: Insights from Twitter Analytics for Policy Makers. 16th Conference on e-Business, e-Services and e-Society (I3E), Nov 2017, Delhi, India. pp.239-250, 10.1007/978-3-319-68557-1_22 . hal-01768515

HAL Id: hal-01768515

<https://hal.inria.fr/hal-01768515>

Submitted on 17 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Deep Analyzing Public Conversations: Insights from Twitter Analytics for Policy Makers

Nimish Joseph ¹[0000-0002-0560-7325], Purva Grover ^[0000-0001-5191-891X], Polaki Kishor Rao ^[0000-0002-7062-7314], and P Vigneswara Ilavarasan ^[0000-0002-9431-3520]

Department of Management Studies, IIT Delhi, India
¹nimishjoseph@gmail.com

Abstract. The paper argues for use of Twitter analytics in the study of public policy. Twitter as a social media platform is widely used by many users to express their emotions on various issues. Governments are now realizing the need for social media for administrative purposes and information proliferation to the general public. However, the extent of knowledge among policy makers, about gathering insights from citizen conversations on various social media platforms needs improvement. This research discusses Twitter analytics for two different use cases namely ‘Demonetization’ and ‘GST’, where citizen engagement and conversations are monitored, to gain insights. These use cases vary in terms of focus, volume, velocity, veracity, time frame, location, and networked-ness. Possible inferences are presented to gain insights and to help in the policy making process.

Keywords: Analytics, Social Media Analytics, Descriptive Analytics, Content Analytics, Network Analytics, Hashtag, Demonetization, GST.

1 Introduction

1.1 Social Media Analytics

Social media is facing an exponential growth and encapsulating almost every citizen across geographies. Social media analytics by definition comprehends filtering of useful insights from the huge data that is available in the form of semi-structured and unstructured public opinion or user generated content [1].

Social media has changed the manner in which we communicate for years now. It is no longer treated as a niche segment for early adopters. Status quo suggests that every possible brand irrespective of size and target audience is capitalizing to make higher profits for better understanding their customers need through continuous public engagement. Gradually, government and public policy makers have realized the importance of social media [2] and are endeavoring to move towards better adoption. Immediate and transparent, social media has provided the public with greater control, participation, and influence over governmental issues and initiatives over the past few years [3]. For example, a pothole on a road leading to transport inconvenience is now directly uploaded on Facebook to draw the attention of the authorities by either making it a matter of mockery or emphasizing the risks that are faced by the public. With the advent of social media, politician/government and policy makers can no more get away with their

false promises as public is actively engaged to evaluate and suggest government through immediate feedback. Surprisingly, contemporary political elections are promoted and run through digital media to get more traction and understand public emotions and demands. While many government agencies still tend to employ the "broadcast" model when using social media, some are engaging through hashtags, community building initiatives, and geo-location analysis [4]. These efforts are helping to better inform the public and alert them to public safety emergencies in real-time. Government organizations often struggle with social media. With legal regulations, privacy concerns, risks in general and lack of knowledge for effective use, policy makers in India are finding it difficult to incorporate social media into their routine decision making formats [5]. A proper guidance from the field experts is required to show policy makers as how social media could be used by them and suggest methods to assess the impact of government social media interactions, but require demonstration on how various tools could be used to capture the insights from online conversations [6]. Therefore, the study shall undertake Twitter analytics for two real issues or cases where citizen engagement or conversations could be monitored and insights can be gathered through various analysis which policy makers can refer to for any future endeavors.

Social Media Analytics is rapidly emerging as an important aspect to drive business [7]. Despite the importance of social media analytics, manuals for easy use by the public officials are not readily available. As the required skillsets are short among the government officials, policy decision making processes seem to miss insights from the social media sources. These platforms can be very well leveraged to understand the people perception about an issue/trend or a policy and utilize the information to revise or amend such policies.

The information sharing can occur in two different ways based upon the insights so generated. Any government while formulating its policies and structures, endeavors to spread awareness and information about the new policies or changes in policies to the intended citizen [8]. Social media can be proven to be a great tool to share the information to an amplified population at a much cheaper cost and effectiveness [9]. One major advantage of social media is that the information sharing is bidirectional i.e. the citizen can respond to these changes [10]. People can post their queries, feedback or complaints directing it to a specific a government department.

Similarly, different government departments can interact with each other for general information or campaigning [11]. This can also be extended to international boundaries as well. For instance, Department of External Affairs interacts with different global governments to resolve issues possibly every other second.

In this study, we are analyzing two most recent and relevant topics through hashtag analysis. The cases studied in this paper are-Demonetization and GST.

Demonetization: On 8 November 2016, the Government of India announced the demonetization of all ₹500 and ₹1,000 banknotes which means that they ceased to be legal tender [12]. The government claimed that the action would curtail the shadow economy and crack down on the use of illicit and counterfeit cash to fund illegal activity and terrorism [13].

GST: Goods and Service Tax is an attempt to unify all taxation of goods and services to provide a seamless tax system for the entire country. The possible advantages of GST are that it will remove multiple taxation system to create a single market [14].

In our study, we have used Twitter, the fastest emerging social media network, as a social media platform for our analysis. The tweet count for Demonetization is 159433 and GST is 33570 tweets respectively. Comparison of the two issues depicts significant differences in terms of focus, volume, velocity and time frame. The data has been collected from November, 2016 to February, 2017. Figure 1 and 2 give the details of tweets collected.

Demonetization addresses and outreaches almost all the population of the country irrespective of the class or group [15]. This is due to the simple fact the money(currency notes) are used by everyone almost every day for transactions and therefore, any manipulation in such policies would affect the whole country. On the other hand, subject like GST, interests economists, analytics and businesses more when compared to the general population. Even though the issue scales at national level but not everyone is well aware about it and even if they are not many tend to express their views on social media.

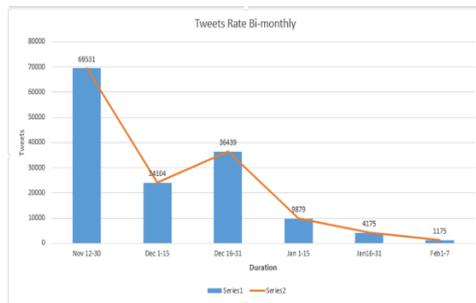


Fig. 1. Tweet rate for Demonetization

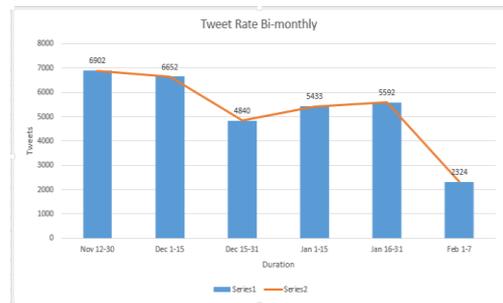


Fig. 2. Tweet rate for GST

From the above 2 figures it is very much evident that demonetization is more impulsive in nature considering the fact that the tweet rate steeply decreased from 69531 in the initial 15 days of announcement to 1175 by the starting week of Feb, 2107. On the other hand, tweet rate for GST is more or less consistent and continuous over the weeks of data collection period. These graphs help to understand the scale of the discussion and the engagement rate of the people during that time period. Demonetization has clearly higher scale and people are discussing about it in the initial phases but the engagement is fading away gradually over a period of time. GST, the scale is low but the engagement is fairly constant for the period. The probable reason might be the fact that GST is yet to be implemented so people are not well aware of what consequences it might have after it is practically implemented. So, people are discussing more generically based on their expectations and knowledge about the subject.

2 Research Method

The analysis of the tweets has been done using R and NodeXL. Twitter API was used to extract the data. The study was divided into three major stages [16].

- **Data Collection:** Collection of data using relevant hashtags, i.e. tweets with #Demonetization and #GST were extracted.
- **Data Cleansing:** This involved cleansing of tweets as it is important from the analysis standpoint. The extracted tweets were first converted to lowercase text, post which the hashtags and hyperlinks were removed, along with punctuations, stop words and chains, retweets, special characters and @usernames etc.
- **Data Analysis:** The filtered data was then used for the analysis. Subsequent section discuss about the various analysis performed.

3 Analysis

Three different types of analysis were performed to extract insights from the data.

3.1 Descriptive Analytics

Tweet statistics. The total number of tweets collected for count for Demonetization is 159433 and GST is 33570. For demonetization the tweets contained #Demonetization or #Demonetisation and each such tweet most probably included multiple related hashtags. Similarly, for GST, all the tweets included #GST and most probably had multiple related hashtags.

Inference for policy makers. These data allows the policy makers to comprehend the scope and scale of the issue. Demonetization with total tweet count of 159433 when compared to that of GST with 33570 tweets is a clear demonstration of the fact that demonetization impacts a larger audience and therefore has a larger scale of engagement. This data can also be used for prioritization of things from the task list. If the engagement is larger, more people will participate in the discussion and therefore there is higher probability of noise. Under such scenario, the authorities can control the noise and ensure right information is spread. With limited resources and infrastructure available, government can direct and concentrate more towards topics with higher engagement.

User statistics. User analysis depicts the number of unique users, their activity, influencing characteristics etc. The details about the user metrics is given in Table 1.

Table 1. User Statistics

	#Demonetization	#GST
Users	61698	12722
Tweets/User	2.584087	2.638736

Unique Tweet/User	1.040974	1.509904
Retweet/user	1.543113	1.128832

Inference for policy makers. Knowledge about issue - The ratios like tweets/user, unique tweets/user, retweet/user give a relative idea about the engagement and spread of the issue but not the absolute values. For example, unique tweets/user is 1.04 and 1.5 for demonetization and GST respectively which shows a comparison that for policy change like demonetization, more people have knowledge about it as more retweets have been generated (60%) whereas knowledge/information about GST is concentrated to only few users and not many users tend to retweet (43%).

User profiling (most visible/ influencers) - Policy makers can get deeper useful insights by understanding the demography of these users. They can be segregated into clusters based upon commonalities and differences. Example, male/female users, who are the influencers, most visible users, location of people posting the tweets, etc.

Influencers by retweets - An influencer can be categorized as anyone with large number of retweets in the network. There are two categories of influencers, some being positive and others negative. Brands target and use these influencers to manipulate decision making of their customers [17]. Similarly, policy makers can utilize such influencers to spread the message to the population in a way that they intended. On the other hand, negative influencers need to consistently monitor to analyze the criticism and respond in a way that minimizes information asymmetry and damage control. This can make the followers to perceive information in opposite way of what actually was intended.

For each of the case under consideration, two of the influencers have been figured out and few of their respective tweets have been analyzed to determine if the users are positive or negative influencers. @YRDeshmukh and @RituRathaur were identified as the top two influencers with respect to #Demonetization and @sachin_rt and @arun-jaitley for GST [18].

Most visible users - By definition, most visible users are the users who tend to participate in the discussion more. These users might not have a great network of followers but they have their visibility in the discussion more often than influencers. Visibility is quantified as number of times a specific user has tweeted within a stipulated time.

Most visible users for Demonetization include user names 'Sreehari', 'mishr sanjay', 'adrenna', etc. and for #GST include 'GST Tracker', 'GlobalSmallTalk', etc [18].

Inference for policy makers: These users tend to take part in the discussion more often and therefore probability of their availability over a certain time frame is more. Similarly, policy makers can optimize the selection of most visible influencer to outspread information to the public in a quick and timely manner. From the above case tweets made by the user suggest their stance towards the issue which needs to be understood to categorize them into positive and negative most visible users.

- SHREEHARI- #Demonetization is solving unexpected problems. no funding for stone pelters and maoists! (Positive)
- Mishr_sanjay- @manjari__ @iSirArnab india knows that only BJP mercenaries are spreading these lies. after @MamataOfficial opposed #demonetization (Negative)

- GST_TRACKER- GST to create Rs. 36,000 crore software market in MSME segment.
- Globalsmalltalk: Centre, states gearing up for GST, says CBEC member (Information and statistics- Neutral)

3.2 Content Analytics

Most frequent Hashtags.

Figure 3 gives the most frequently hashtags for both Demonetization and GST.

#Demonetization		#GST	
#Demonetization	45000	#GST	18726
#Demonitisation	15256	#GSTbill	1543
#BlackMoney	1790	#DeMonetisation	1272
#India	1108	#GSTCouncil	857
#Modi	1032	#Budget	720
#cashless	564	#tax	499
#IAmWithModi	510	#GSTNews	423
#BJP	449	#GSTIndiaExpert	401
#CashlessEconomy	441	#ArunJaitley	397
#RBI	384	#RunSimple	391

Fig. 3. Most frequent hashtags.

Inference for policy makers. Most frequent hashtags [19] allows policy makers to understand the trend among the citizen about a particular issue or policy. They can determine the most frequent words that are there in a discussion and are attached along with a specific issue or policy. The inclusion of these words will make the information reach the target population more effectively and efficiently. For example, if there is any new policy amendment that the government has made in the existing ones and it wants to disseminate the information to general/target population then it can refer to the most commonly used hashtags about the particular issue and post tweets including some of these hashtags to have an amplified outreach.

From the above data, one interesting information that can be gathered is that both the topics have Demonetisation as one of the top frequent words.

Few of tweets are:

- @ShekharGupta @ndtv Would be real shame if GST gets delayed because of demonetisation. Too much pain and very little gain so far #GST
- #Demonetisation, #GST to transform India's business ..
- can #Country suffer double #destabilisation of #Demonetisation/ #GST at one go - the #GST should wait till States economy improves #Mitra
- #GST is good economics, #demonetisation is not.

We can see that in this particular tweet the user is directing his concern to authorities by stating that there is no cash available in ATM. Government can use such information for resolving the concerns of the citizen or giving them an update.

- #Demonetization has come as bliss for indian #PSU banks which have had been marred by #NPA . #SBI has got 35% of total cash deposits

This tweet depicts that the banks have been happy about the policy changes as it has led to reduction of NPAs.

If we also extract tweets for hospitals, the following tweets were found:

- @HospitalsApollo when whole India is moving towards cashless economy your doctors are asking for cash #Demonetization

Clearly, citizen are facing problem with healthcare services at Apollo hospitals. People are not denied treatment because of lack of cash. Government needs to address such issues by monitoring, interpreting and responding immediately to prevent public pain. Government can use such analysis for emergency management services and mitigate potential crisis, social unrest can be managed effectively through active listening, identifying noises and false information, and taking action to defuse or suppress them, such as issuing clarifying statements as necessary to dispel rumors.

Sentiment Analysis.

For sentimental analysis sentiment package for R has been used for classifying tweets by emotion, and polarity.

Negative polarity is defined as for any tweet that has negative emotion like sadness, anger, fear or disgust. Positive polarity is for the tweets including emotions like joy and trust. For, neutral emotions like surprise and anticipation tweets are classified as neutral. P-Positive, N-Negative, NTL-Neutral respectively.

Sentiment Analysis by polarity (words): The overall sentiment for each case has been graphically represented using Figure 6.



Fig. 6. Sentiment by words polarity.

Sentiment Analysis by polarity (tweets): The overall sentiment for each case has been graphically represented using Figure 7.

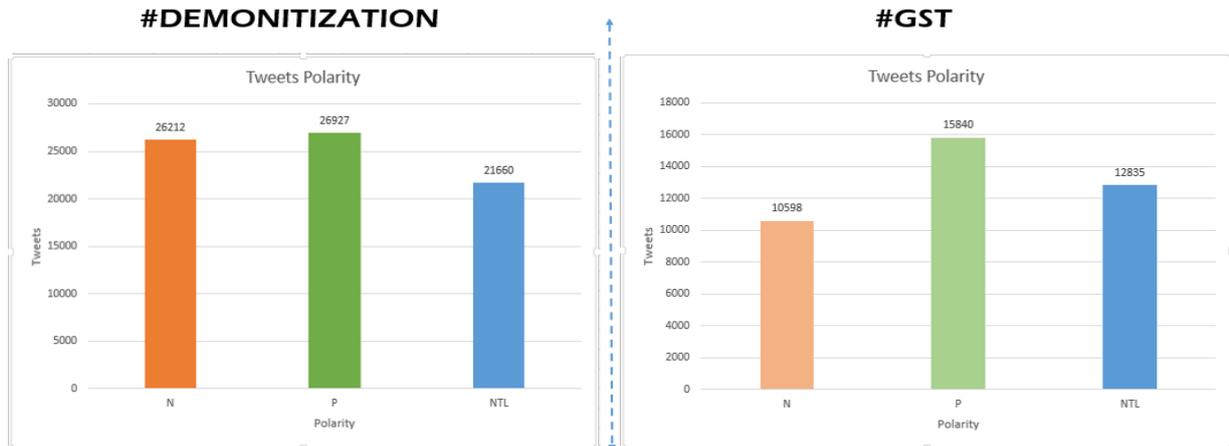


Fig. 7. Sentiment by tweet polarity.

Demonetization.

- Negative tweets- 35.07%
- Positive tweets- 36%
- Neutral tweets- 28.9

Inference. Positive and negative tweet count by percentage is almost same suggesting that there is an ambiguity and people are not very much satisfied with the policy change.

GST.

- Negative tweets- 26.98 %
- Positive tweets- 40.33 %
- Neutral tweets- 32.69 %

Inference. Positive tweet percentage is 40.33% whereas negative tweet count by percentage is 26.98% suggesting that there is overall positive emotions among the people about the policy change/implementation.

Sentiment by emotion for both the cases are depicted using Figure 8

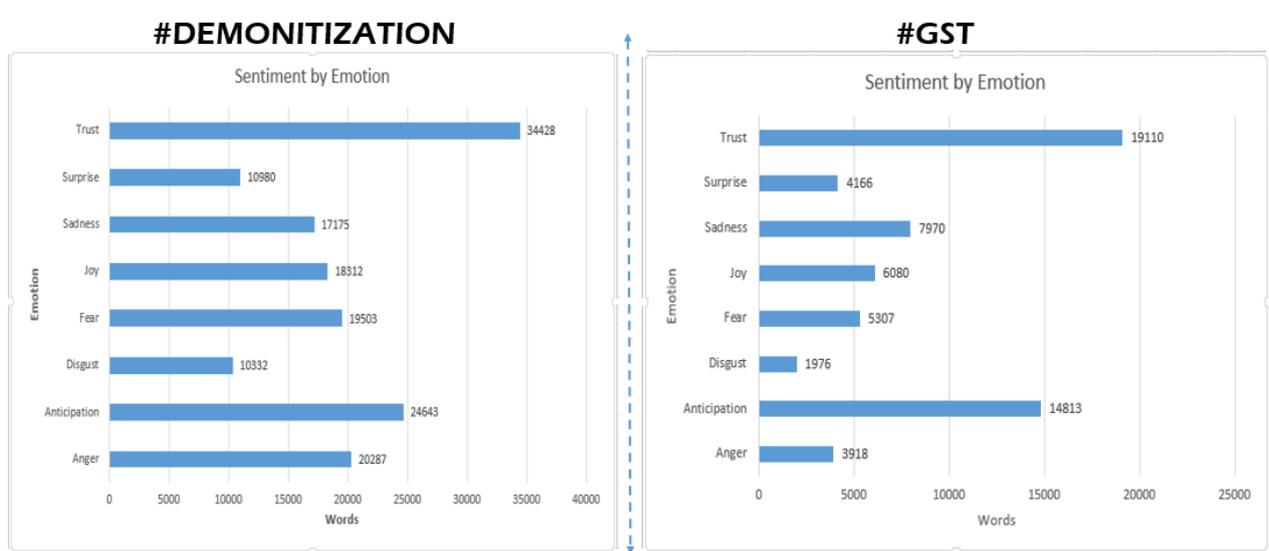


Fig. 8. Sentiment Analysis by emotion.

From both the graphs two specific emotions are reciprocated very clearly. Trust and anticipation word count are more when compared with other emotions. Usage of more trust words suggests that people are supportive of the policy. But, is not a true measure because, considering tweet polarity, negative tweets are more in number.

Demonetization: Trust + Anticipation (34428+24643) words

GST: Trust + Anticipation (19110+ 14813) words

Inference for policy makers. Sentiment analysis is one of the most effective way to gauge public response of a policy by the government. Understanding emotions of a tweet/content helps policy makers to classify people into groups and take actions according to their responses on an issue. Figure 10 explains how sentiment analysis could be used.

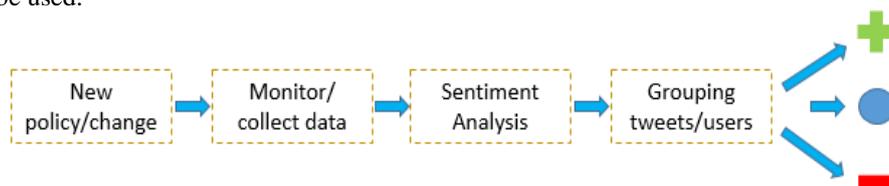


Fig. 9. Sentiment Analysis

Some of the questions that can be very effectively answered through such analysis are:

- What do citizens feel about the new policies and initiatives?
- What are the most talked about the new policies/amendments?
- What are the most positively talked about attributes in the new policies?
- Who are advocates and skeptics of the new policies?
- Where the government should be actively listening?

3.3 Network Analytics

Topological Analysis

A social network of the influencers was created to understand their network and relationship. The network include followers, followers of , etc. Lay out was done by Fruchterman & Reingold representation. Influencers like @YRDeshmukh and their network was identified for #Demonetization and the influencers in the #GST network identified are BJ4India, Sachin_Kingdom, Bambamroy1980.

Inference for policy makers. Government can use such analysis to determine the largest networks of discussion for an issue and monitor the users to see their behavior and how they are influencing the network and act accordingly by taking corrective measures or using the positive influencers to enhance the reach of the message.

4 Conclusion

With the advent of technology and social media, Government is endeavoring to capitalize on this and use it for improvement in policy making and overall public service. This study is limited to the data from Twitter. However, with more people going online and expressing their opinion (citizen journalism), it becomes more difficult and important to filter out noise form the useful insights that the policy makers can make use of. Hence, it is imperative to build infrastructure and knowledge building in regards with the tools and frameworks. This study explains how government can use social media to improve governance and policy making by enhancing transparency, participation, collaboration and engagement.

Transparency can be achieved by making necessary information available to the public. Governments can use social media platforms to keep its citizens informed about the policies and amendments. Participation of citizen in the discussion allows the government to interpret and comprehend public opinion about the policies. Collaboration of various governmental departments and various public private institutions is possible through the social media. Each department can leverage the power of social media analytics for cross functional insights and better governance. Engagement happens when the information sharing occurs in all the possible direction.

References

1. Omitola, T., Ríos, S. A., & Breslin, J. G: Social semantic web mining. Morgan & Claypool Publishers. (2015).
2. Lee, G., & Kwak, Y. H: An open government maturity model for social media-based public engagement. Government information quarterly, 29(4), 492-503 (2012)
3. Zhang, W., Johnson, T. J., Seltzer, T., & Bichard, S. L: The revolution will be networked: The influence of social networking sites on political attitudes and behavior. Social Science Computer Review, 28(1), 75-92 (2010)

4. Kaplan, A. M., & Haenlein, M: Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, 53(1), 59-68 (2010)
5. Magro, M. J: A review of social media use in e-government. *Administrative Sciences*, 2(2), 148-161 (2012)
6. Hoffman, D. L., & Fodor, M: Can you measure the ROI of your social media marketing?. *MIT Sloan Management Review*, 52(1), 41 (2010).
7. Peters, K., Chen, Y., Kaplan, A. M., Ognibeni, B., & Pauwels, K: Social media metrics—A framework and guidelines for managing social media. *Journal of interactive marketing*, 27(4), 281-298 (2013)
8. Stoneman, P., & Diederer, P: Technology diffusion and public policy. *The Economic Journal*, 104(425), 918-930 (1994).
9. Antcliff, V., Saundry, R., & Stuart, M: Networks and social capital in the UK television industry: The weakness of weak ties. *Human Relations*, 60(2), 371-393 (2007)
10. Java, A., Song, X., Finin, T., & Tseng, B: Why we twitter: understanding microblogging usage and communities. In *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis* (pp. 56-65). ACM (2007)
11. Kavanaugh, A. L., Fox, E. A., Sheetz, S. D., Yang, S., Li, L. T., Shoemaker, D. J., Natsev, A. & Xie, L: Social media use by government: From the routine to the critical. *Government Information Quarterly*, 29(4), 480-491 (2012).
12. Sherline, T. I: Demonetisation as a prelude to complete financial inclusion. *International Education and Research Journal*, 2(12) (2016)
13. Nag, A: Lost due to demonetization. *Economic and Political Weekly*, 51(48), 18-21 (2016)
14. Poddar, S., & Ahmad, E: GST reforms and intergovernmental considerations in India. Ministry of Finance, Government of India (2009)
15. Singh, P., & Singh, V: Impact of Demonetization on Indian Economy. In *3rd international conference on recent innovation in science, technology, management and environment* (2016)
16. Joseph, N., Kar, A. K., Ilavarasan, P. V., & Ganesh, S: Review of discussions on internet of things (IoT): insights from twitter analytics. *Journal of Global Information Management (JGIM)*, 25(2), 38-51 (2017).
17. Zhang, Y., Moe, W. W., & Schweidel, D. A: Modeling the role of message content and influencers in social media rebroadcasting. *International Journal of Research in Marketing*, 34(1), 100-119 (2017)
18. Twitter, <http://www.twitter.com>, last accessed 2017/03/05
19. Stieglitz, S., & Dang-Xuan, L: Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, 29(4), 217-248 (2013).
20. McNaught, C., & Lam, P: Using Wordle as a supplementary research tool. *The qualitative report*, 15(3), 630 (2010)