

Analyzing Tweets Report

Dong Han

CSE 581

Dec. 9, 2014

1. Introduction

This is a report for the Tweets analyzing. Twitter is a Social Network Service, its users can publish messages with number of characters limited in 140. We can consider that twitter is a variation of instant messaging. It allows a user to push text that can be viewed by all users. In twitter, the published texts are tweets. Other users can get their interested tweets by using Twitter service. By using twitter, users can share their information more convenient. If a user likes a tweet, the tweet can retweet the tweet. Then a visitor can view the tweet by visiting the page of the user. We can consider that twitter is a public platform, you share what you see, what you think, and what you care about. You can also browse the content from other users in real time and join common topics. In traditional media, there are professional editors or announcer to broadcast news. In the world of twitter, everyone can publish tweets to talk about some new events, which is faster than traditional media. In twitter, if you care about some users, you can follow them. When they have new status, the updated information will be pushed to your twitter application. Nowadays, the social media likes twitter is developing very fast. Users can view tweets by web or mobile devices and spread their tweets in their social circles.

Hashtag is a feature that is widely used in twitter service. Hashtag is a topic label to refer the content of the tweet that is created by publisher. Hashtag can help users to find hot topics from massive messages in Twitter. Because any user can arbitrary add hashtag in their messages. Usually, the same hash tag may refer to different topics. Though many tweets use a same hashtag. These message may also discuss different subtopics of a main topic. Mining the different topics which are using a same hashtag has significant meaning, because it can help us to discover new topics and present messages to different topics. For the social media services like twitter, the information are mainly created and broadcasted by general users. On the one hand, massive messages and topics are published. On the other hand, because these messages do not have strict structure [1]. When we review these messages, it is not hard to find out that these messages are fragment text and many of them are redundancy. In Twitter, hashtag is described by letters and numbers, with “#” as its prefix. By using hashtag, a user can quickly find an interesting topic. For example, when a user click a hashtag “#apple”, all messages relate to the apple are retrieved. Hashtag is a good feature to show tweets for some topics. We cannot simply consider tweets with some hashtag can be classified to an exactly same topic. For tweets about a hot topic, they share a same hashtag, but they may talk about different things that relate to the topic.

In the project, I will study how to find subtopics from a hashtag [2]. In twitter, if you do search for a hashtag, you can get many messages that include the hashtag, as shown in Figure 1. Though these tweets relate to the hashtag “#apple”, they can talk about different things in the area. For example, in Figure 2, I analyze some tweets from a search result manually. You can find out that these tweets belong to three subtopics. However, if we search a hashtag but get massive results, we should use an automatic method that can help use analyzing the subtopics in the results. By cluster the result to a few subtopics, we can focus more on the tweets that we are interested in without viewing results for unrelated or redundant tweets. In addition, a tweet is short text. If some tweets are discussing the same subtopic, we also have difficulties to put them in a same subtopic, because they may not use identical words to describe a same event. This is also a problem that I will study in the project.

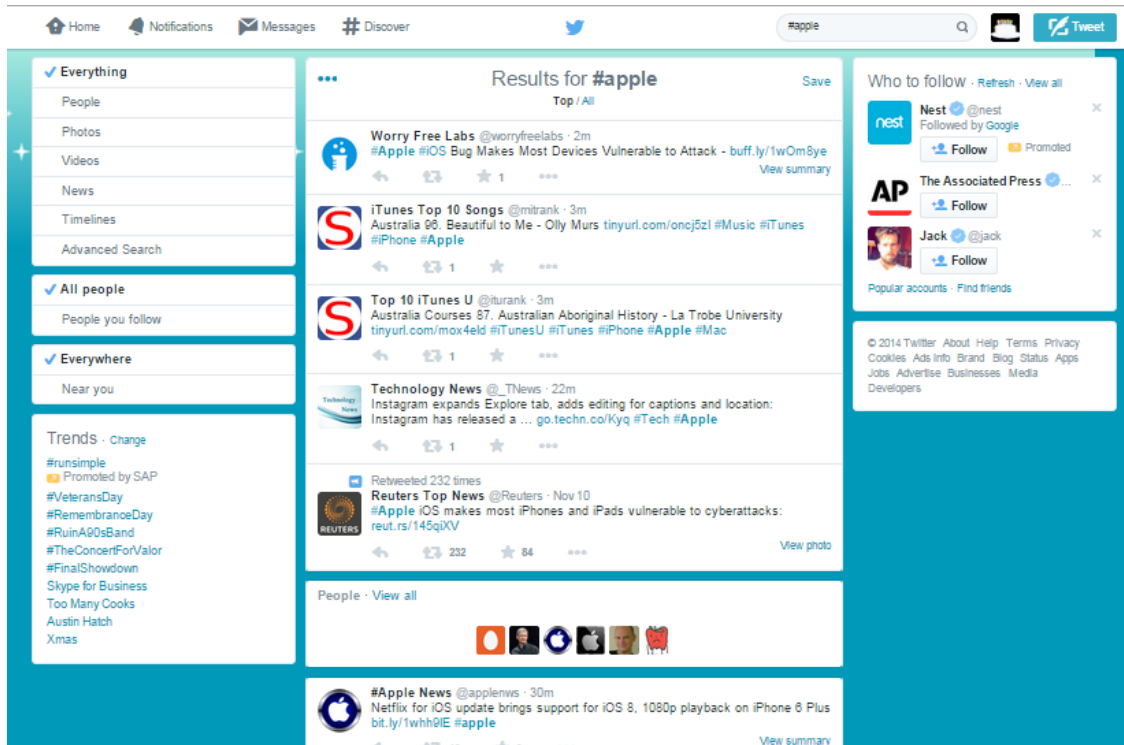


Figure 1 the result of searching hashtag “#apple”

Example: a search result for #apple

Designer News Bot @designernewsbot · 6h
Apple Patents An iPhone Drop Protection Mechanism That Changes Device Angle In Freefall [#apple news.layervault.com/click/stories/...](#)

Achieve @AchieveControl · 43s
[ift.tt/1w1BVGg](#) Review: Skech Crystal for iPhone 6 and iPhone 6 Plus, Hard Rubber + Ice for iPhone 6 - iLounge [#Apple](#) [#Tech](#)

George Ch @George_Cha · 7h
[#Pebble](#): a wearable [#Apple's](#) Watch doesn't see coming. Or they do but they exist in a different profit bracket. [zite.to/1vLAjTu](#)

OneAppleWorld @OneAppleWorld · 1m
UBS Survey Predicts Big Consumer Appetite in 2015 for Apple Watch - Tech Times [bit.ly/1AcDZii](#) [#apple](#), [#watch](#)

iOS Jailbreak @GreenPois0n_JB · 8h
Apple Stops Signing iOS 8.1, Downgrade Is Impossible [ow.ly/FdjCE](#) [#Apple](#) [#Jailbreak](#)

#Apple News @applenws · 5h
Dropbox adds file/folder renaming and Office document editing to iOS app [aol.it/1AbGTDU](#) [#apple](#)

Cluster the tweets to subtopics

- iPhone devices features
- Apple's watch
- iOS features

Figure 2 an example of subtopics

2. Motivation

When we search a hashtag in twitter, it just return results that contain the hashtag. In these results, they may refer different subtopics that relate to the hashtag, or they may concern other topics, because of the various meaning of one hashtag. In twitter, it does not give further analysis about how many subtopics in the search result. For a user, to find interesting subtopic is an important feature in searching. In the project, I want to collect tweets from hashtags search result, then cluster a search result to appropriate subtopics, which can help user to discover interesting and valuable topics in massive tweets.

3. Method

3.1 Overview

The whole workflow in the project is from data collection to clustering. There are several issues we need to study. Now each step is briefly described. A general workflow is shown in Figure 3. A detailed description of each step and related issues is discussed in following subsections. At the beginning, we should get search result of a hashtag from Twitter. If we do it in webpage of Twitter, we just need to type in the hashtag, then submit the search request. The Twitter server will respond latest results for the keyword. However, the Twitter server does not return all results relating to the topic. To automatically get search result as many as possible, we use Twitter API that is provided for Twitter third part application development. After we get the raw data from the search result, we consider to process the raw data in the second step to pre-process the raw data we just get from previous step.

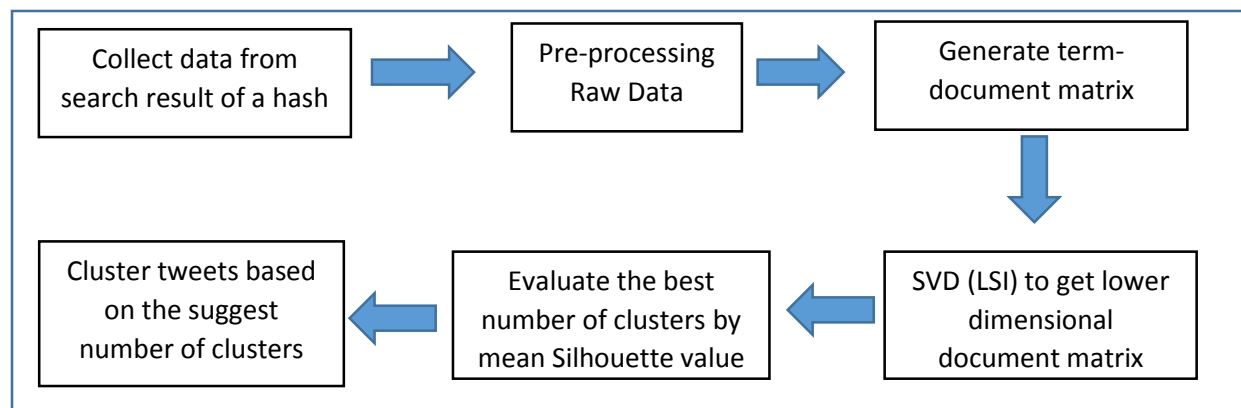


Figure 3 workflow of data processing and clustering

In the second step, we tokenize each tweet, remove stop words and do some other processing. Then we output the cleaned data to the next step for term-document matrix generation. Because we have a large number of tweets and tokens. To improve the efficiency in weighting each token in each document, we use sparse matrix. Another reason to use sparse matrix is the structure can save a lot of storage space. Then we put the term-document matrix to a SVD solver. The SVD solver processes the matrix and outputs a transformed document matrix with reduced dimensions. Then we do clustering on the transformed document. We consider that an optimal number of cluster in a predefined interval. Then

we use Silhouette value to evaluate the clustering result for each clustering set up and choose the one has the best result in Silhouette value. After these steps, we get to the final step, where we use the adjusted number of clusters to cluster the transformed documents to subtopics.

3.2 Data Collection

Twitter's API [3] is opened as JSON based web service. If you connect to a Twitter server and request a service by following JSON document format with specific parameters, you can get response from Twitter server. The responded data is also a JSON formatted file, with key-value pairs. To avoid abuse, in a pre-defined period, you can just submit limited number of request. For example, for searching request, you can only invoke searching API no more than 150 times in 15 minutes. In addition, the Twitter service cannot be anonymously accessed. To build an application that can grab search result from Twitter, your application should be authorized by Twitter. In Twitter service management site, you can apply application keys, which is used for authentication. When the application passed the verification, your application is granted to access Twitter's server by using Twitter API. To make the Twitter API more easily using, some third part library wrap the Twitter API. These third part library can help developers to assess Twitter API by using many programming languages. In the project, I choose Twitte4J library, which is designed for Java language. Twitter4J wrapped lower lever of Twitter API and gave a more convent interface to developers. Twitter4J can help you to encode your request to JSON formant and then decode the JSON response to many variables in Java. To get the response result, we just need to check values in different variables. In the current project, to simplify the data processing, in searching query, we add a constraint that the retrieved tweets are using English.

After using Twitter4J [4] to get raw result of searching, we still need to process the data to make it more easily to be used in term-document generation. For the steps in data pre-processing, you can refer Figure 4.

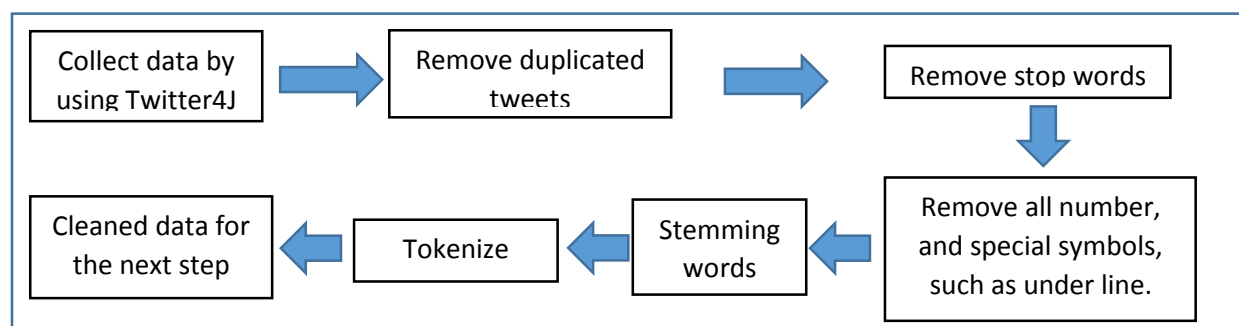


Figure 4 workflow of data pre-processing

For the duplicated tweets removing, we add tweets as an entry in a table. If there is a new tweet to be added, we compare the new tweet with existing tweets in the table. If there is no existing tweet that is the same as the new tweet, we add the new tweet as an entry of the table, or else we discard the tweet.

For the data cleaning, we use Lucene library [5] to do the proccession. Lucene library is a third part library developed for java, the library is used for nature language processing and searching. In the project, we use some basic functions in Lucene library for text tokenization. First we add a stop words list to the

Lucene processing model. After all stop words are filtered, we use regular expression to remove numbers, and special symbols, such as underline. Then we use a Lucene model to stem words and tokenize words. After doing this, each tweet is divided into a sequence of tokens. After we convert each tweet to sequence of tokens, we can build term-document matrix.

3.3 Term-document matrix generation

To compare the similarity of documents, the vector space model is proposed. In the model, the documents and terms are weighted. For each document D , it is converted to a vector, $D = \{d_{i1}, d_{i2}, \dots, d_{in}\}$. The each element in D means the weight of the term in the document. If we consider each term as a dimension, then they compose a space with n dimensions. In the document set, each document can be converted to the vector. This we get term-document matrix. As shown in Figure 5, each document can be shown in the term space. For the similarity of two documents

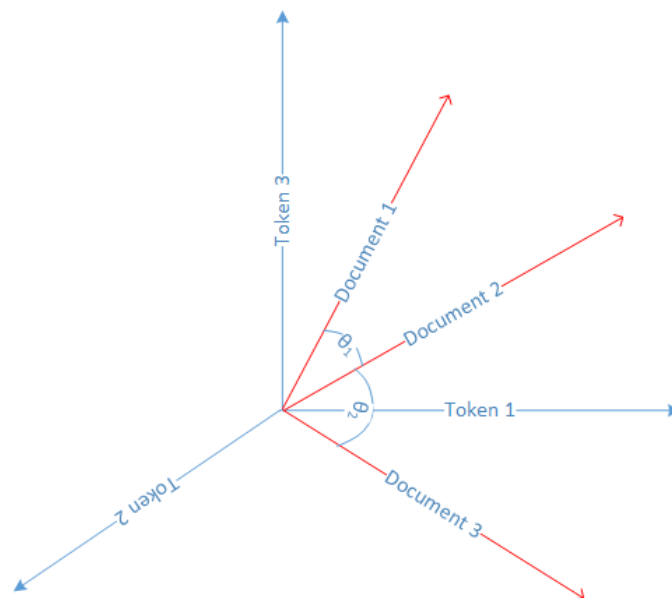


Figure 5 documents in tokens space

The advantage of vector space model is that it can convert unstructured document to vector, then we can use mathematical tools to process the document. In traditional vector space model, because the over simplified model, we consider that all tokens do not have any correlation. In the document space as shown in Figure 5, each pair of base vectors are orthogonal. However, in a real case, the assumption of orthogonal is not established. In natural language processing area, one word usually has multiple concepts, or we use different words to describe one concept. Moreover, the semantic structure between words are implied in the relations of words' frequency. By using statistical method, we can eliminate the influence from one word has multiple meanings, and multiple words specify one meaning. Latent semantic Indexing is one method that can solve the issue.

To use traditional matrix to store term-document matrix is storage consuming. For example, if we build a term-document for hashtag "#apple", the search results of "#apple" have 36,715 terms and 67,899 tweets. Then the scale of the term-document matrix is $36715 * 67899 = 2,492,911,785$. If we use 32bits

(4bytes) integer type to store one element, it needs approximately 10GB storage space just to store the matrix. However, because each tweet has no more than 140 character, therefore, for the vector of each tweet, it has many zero elements. We can consider the term-document matrix is a sparse matrix, i.e., the density of the matrix is low.

Weighting calculation is important in building term-document matrix. In the project, we use tf-idf weighting method. We define TF of a token in a document as ratio of number of times a token in a document to the total number of words in the document. IDF of a token is logarithm of the total number of documents to the number of documents have the token. The tf-idf is production of tf and idf.

3.4 Latent semantic indexing

The latent semantic indexing [6] method is widely used in query. In a search engine, there are many documents. To find documents that match your query is a difficult question, especially for the semantic meaning of documents match your query, but not exact words match the query string. LSI can map each document from the original vector space to a lower dimensional space. In the new dimensional space, though two documents do not use any same keywords, they can show similarity in the space.

Singular valued decomposition (SVD) is a general method to construct LSI space. For any $m \times n$ term-document matrix, it can be decomposed to production of three matrixes.

$$X = U S V^T$$

In the expression, each column of U and V are orthogonal. The norm of each column is unit length. S is a diagonal matrix. The elements on diagonal are singular values, with descending order, i.e., $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. In the matrix X , m is the number of rows, n is the number of columns. We denote r and the rank of matrix, that satisfies $r \leq \min(m, n)$. The decomposition is named the SVD of X . Figure 6 is an example of SVD operation.

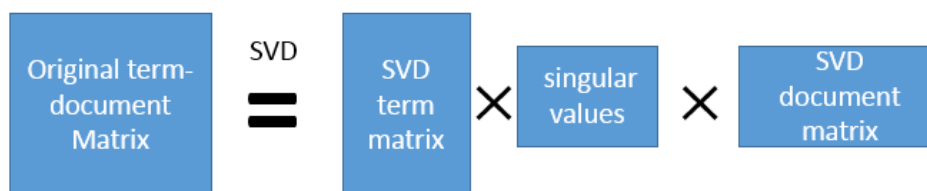


Figure 6 SVD operation

In matrix S , the diagonal is ordered by their importance. We can keep the largest k elements and discard the smaller singular values. At the same time, we remove some columns at the right side of U and V . Then we get a new matrix $X' = U_k S_k V_k^T$. By mapping the original documents to the k dimensional vectors, we can get X' , which has the smallest square error when comparing with X . It means that the LSI method is the optimal dimensional reducing method. When we compare the similarity of two documents, we convert them to the reduced dimensional space. Then we can compare their cosine similarity in the k dimensional space. In the decomposition, the original matrix is transformed to the SVD document matrix, which can be considered as the transformed matrix with reduced dimensions.

SVD is an eigenvector extraction method. It ignores unimportant features of word and keep main semantic information of document set.

3.5 Using Silhouette value to evaluate clustering

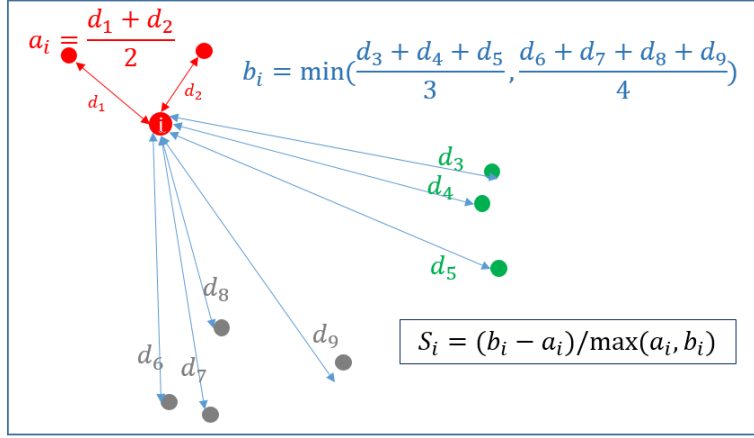


Figure 7 An example of Silhouette value

The Silhouette value of a document is used to evaluate the ratio of similarity of the document with other documents in the same cluster to the documents in other clusters. The value is in the interval [-1, 1]. The formula for Silhouette value is shown in the formula.

$$S(i) = \frac{\min(B(i, C_j)) - W(i)}{\max[\min(B(i, C_j)), w(i)]}$$

In the formula, $W(i)$ is the average distance from instance i to the other instances of a same cluster. $B(i, C_j)$ is the minimal average instance from the instance to the other instances in a different cluster. Figure 7 is an example of silhouette value for point i . In the example, there are 3 clusters. Point i belongs to the red cluster, which has 3 points. The green cluster has 3 points, the grey cluster has 4 points. In the red cluster, the average distance from i to other points are calculated in a_i , for the different clusters, we calculate the average distance from i to all points in green cluster and all points in grey cluster, respectively. Then we set b_i as the minimal average distance from i to other clusters. Then we get S_i as the Silhouette value of point i .

In the project, we evaluate the clusters by using mean value of Silhouette value. If we have n points, then we calculate Silhouette value from point 1 to n individually, that is S_1, S_2, \dots, S_n . Then the mean Silhouette value $\text{mean}(S) = (1/n)(S_1 + S_2 + \dots + S_n)$ is used to evaluate the quality of a clustering.

Because we have massive documents for clustering, the time complexity of Silhouette value calculation is high, therefore, we consider sampling the points, then calculate the mean Silhouette value in the selected samples. We do the procedure many time, then use the average of mean Silhouette value as the final number of clusters.

4. Experiment

4.1 Collected data analyzing

Searched hash tag	Number of tweets	Number of unique tweets	Ratio of unique tweets to number of tweets	Number of tokens
#apple	134,173	67,899	50.60%	36,715
#google	91,093	52,348	57.47%	32,638
#microsoft	44,322	20,037	45.21%	13,828
#michigan	14,370	9,644	67.11%	13,151
#cometlanding	103,697	94,669	91.29%	32,663

Figure 8 a table to show some basic properties of the collect data.

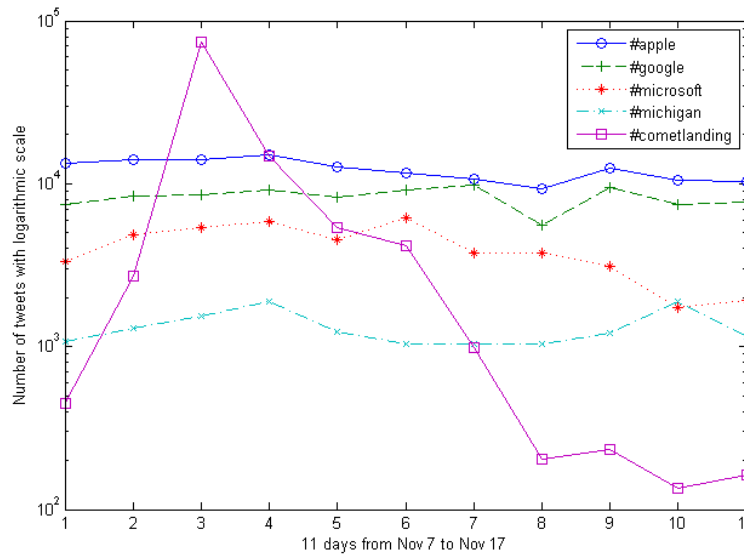


Figure 9 the number of collected data in each day. Y axis is logarithmic coordinate.

We collect results from five different hash tags for 11 days, in the period from Nov. 7, 2014 to Nov. 17, 2014. The basic statistics of results are shown in Figure 8 and Figure 9. In Figure 8, one interesting metric is ratio of unique tweets to number of total tweets. From the metric, we can get redundancy of results for each hashtag. The higher the ratio, the more diversity in results. In the hashtag “#microsoft”, it has the lowest unique ratio. The reason is that, in the result, there are many tweets are created by apps from Microsoft platforms. If many users use some apps, then the app on the devices of these users can automatically post same tweets for the app’s promotion and advertising. From Figure 9, we can find out that the number of tweets for “#cometlanding” increasing fast for the first few days, then retreat fast after the peak at the third day. When we began to collect the result included “#cometlanding”, the Philae lander was just released to comet from Rosetta orbiter. Therefore, there were many people

focused on the event and posted many tweets. After a few days, the landing was successfully finished, then the number of tweets talking about the topic was significantly decreased.

4.2 The efficiency of SVD

Searched hash tag	Density
#apple	0.0249%
#google	0.0239%
#microsoft	0.0565%
#michigan	0.0609%
#cometlanding	0.0220%

Figure 10 the density of each dataset.

Because a tweet's number of characters is limited in 140. We can conclude that the term-document matrix for tweets are sparse matrix. To verify the conclusion, we calculate the density of each search result dataset as shown in Figure 10. The density here is defined as the ratio of number of non-zero elements to the number of zero elements in term-document matrix. From Figure 10, we can find out the density is really low in the dataset.

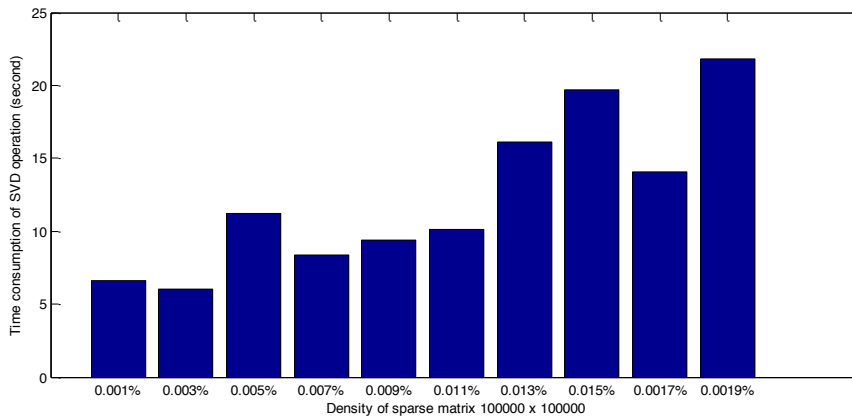


Figure 11 the running time of SVD for different density of sparse matrix. The sparse matrix is 100000 x 100000, and densities are from 0.001% to 0.0019%.

Then we do an experiment to check the efficiency of SVD for sparse matrix. The SVD operation is time consuming. If the scale of dataset is large, it is not possible to use SVD to process the whole dataset. However, for the sparse matrix, we surprised find that the SVD is efficient. In Figure 11, we do SVD for a sparse matrix with density from 0.001% to 0.0019%. The scale of the matrix is 100000 x 100000. In the experiment, a random 100000x100000 sparse matrix is generated. The scale of the matrix is near to the scale of term-document matrix of tweets. We initialize the matrix with specified density. Then we do SVD for the matrix. We run the SVD operation in Matlab on a computer equips AMD FX-9590 Eight-Core 4.71 GHz Processor. The experiment is ran on the Windows 7, which has multiple background tasks running. To avoid the influence from other tasks, we run the experiment many times, and choose the

best result for each density set up. The best result is the one using the minimal time. The result shows that the scale likes the tweets' term-document matrix can apply the SVD and get solved with in tolerable time (less than 25 seconds).

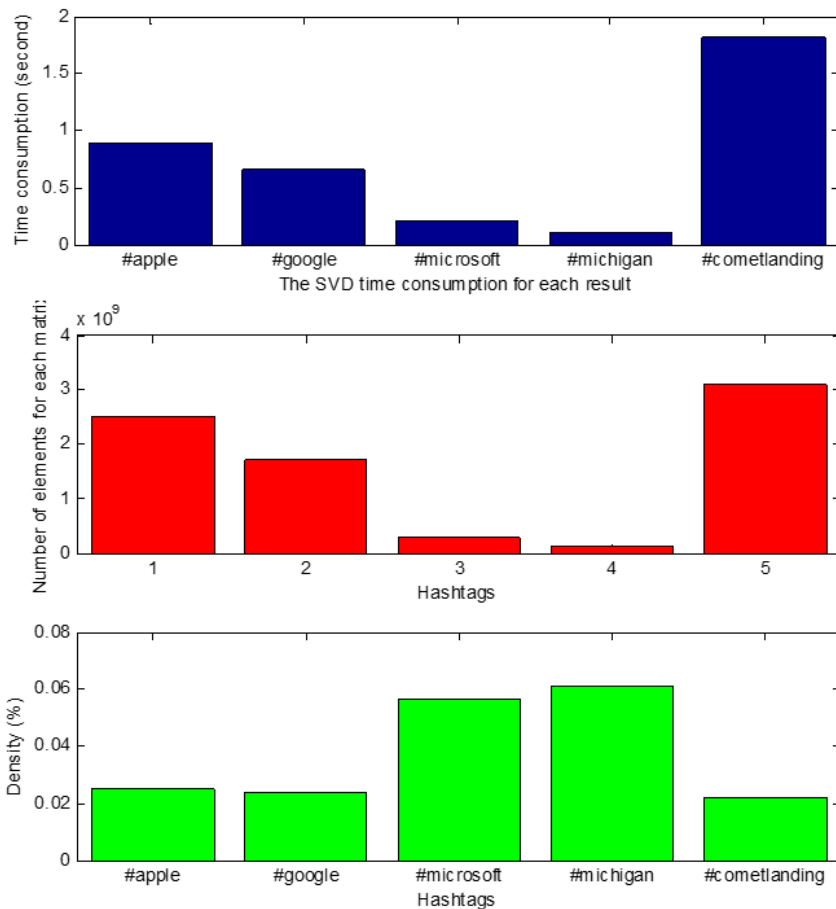


Figure 12 The SVD operation for tweets' dataset, the top figure shows the SVD operation time need for each dataset. The middle figure shows the number of elements for each term-document matrix. The bottom figure shows the density of matrix.

The SVD operation on the sparse term-document matrix is fast, as shown in Figure 12. From Figure 12, we can see that the time needed for SVD highly correlate to the number of elements in a matrix.

4.3 The result of LSI

After the SVD, we get SVD document matrix. The matrix is the transformed term-document matrix. In Matlab, the SVD on sparse matrix can keep the largest six singular values. Therefore, the transformed document is determined by a vector with six dimensions. We draw plot for each dataset with the first two dimensions, the result is shown in Figure 13. Then I will do cluster based on the transformed matrix. However, the number of clusters is difficult to be determined. We begin the clustering from the number of 1 cluster to number of 20 clusters. We evaluate the quality of clustering result after we finish a clustering operation. We can use any clustering method, such as k-means or hierarchical clustering. In

the project, I choose k-means clustering. We begin the clustering from k=1, then we do the clustering for k=2, and k=3, ... , k=20. For each specific k.

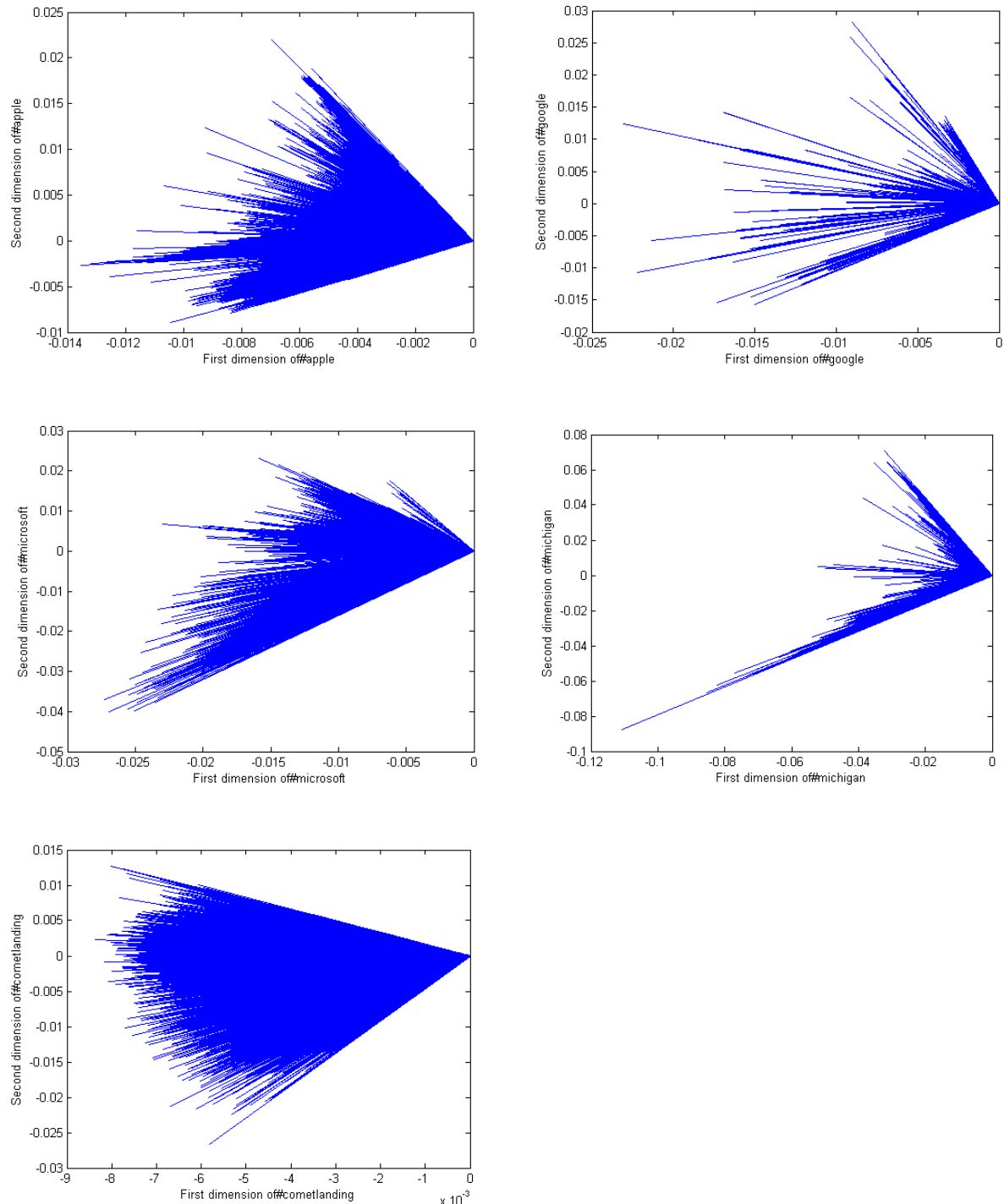


Figure 13 the first two dimensions of transformed documents in LSI analysis.

4.4 Clustering evaluation

We use the dataset of #apple as an example. Because the number of tweets is huge in the dataset, we cannot directly use Silhouette value to calculate each instance in the dataset. We use sampling method to calculate Silhouette value. After a clustering, we get k clusters. Then we randomly choose 2000 samples from the dataset and calculate their mean Silhouette value. We do the sampling for 100 times. Then we get the average mean of Silhouette value for the k clusters. For the dataset of #apple, we get a result as Figure 14. The Figure shows that when we set k=7 for the k-means method, we can get the best clustering result in the interval k=1 to 20. I attached the diagram of mean Silhouette value of other datasets in Appendix.

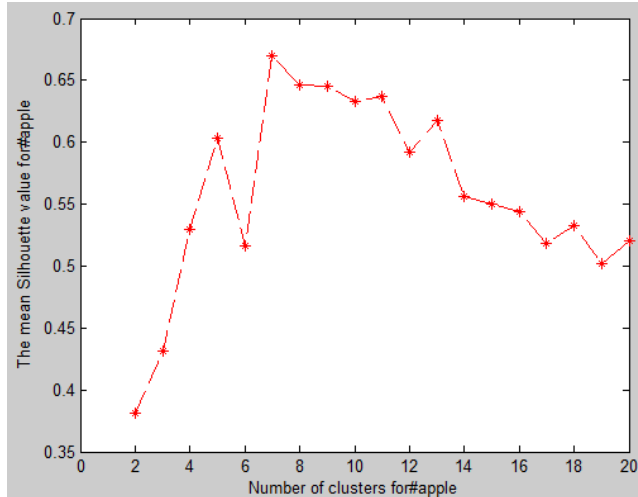


Figure 14 the mean Silhouette value for each number of clusters.

4.5 Clustering result analysis

We show the clusters on the 2-dimensional figure as in Figure 15.

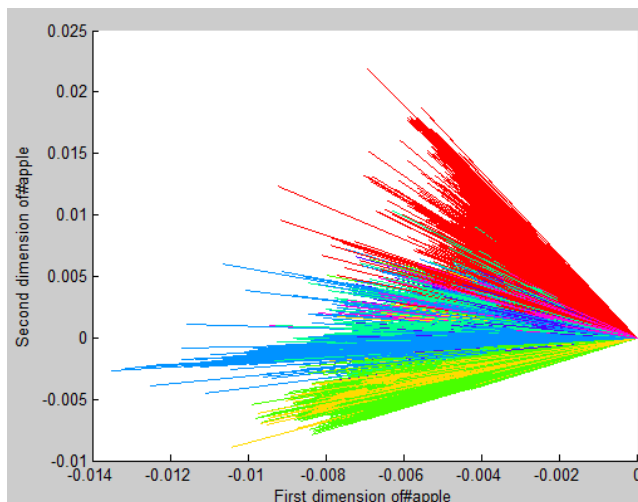


Figure 15 #apple clustering, 7 colors for 7 different cluster.

From the Figure 15, we can find out some cluster are mixed together. Because we are clustering on six dimensional space. Here, we just show the major two dimensional space. Some clusters are discriminable in the rest of dimensions. The clustering result for the other 4 hashtags are attached in Appendix.

For each group, I select two tweets as representatives of the group. It is not hard to find that each group has its main subtopic.

- In group 1, it talks about some prices related news.
- In group 2, tweets are about some things from different countries.
- In group 3, it is about entertainment news.
- In group 4, tweets are discussing ios features.
- In group 5, some news are about cover protector of iPhone.
- In group 6, it is about “smile” as its topic.
- In group 7, some user experience about apple products are posted.

From the analyzing, we know that the tweets in a same cluster have a similar topic.

Group 1:

37415: hertz: we actually made \$ million less than we thought (htz) #news #phone #apple #mobile

17174: oil prices are falling #news #phone #apple #mobile

Group 2:

7462: spain collections . intensive english s - sit tv production unit #itunesu #itunes #iphone #apple #mac

5528: china courses . stage japanese - kolbe catholic college #itunesu #itunes #iphone #apple #mac

Group 3:

28937: united states game paid . dude perfect - dude perfect error #itunes #iphone #apps #apple

67647: the game (season). hd premiere now on itunes. [\$.] #apple

Group 4:

28660: #ios #jailbreak now compatible with macs #apple #ios #tech #news

64760: top #ios news: no #iphone plus recall office for iphone #apple watch to be big bucks

Group 5:

67013: #apple #iphone - hybrid id credit card holder case cover with stand for apple iphone plus bl... #deals ebay ca

56013: #apple #iphone - green . mm super thin matte transparent protective case cover for iphone : ... #deals ebay uk

Group 6:

10899: new #apple #macbook - item - smile if you love an arts administration. sleeves for macbooks

44368: new #apple #macbook - item - smile if you love a gerontological nurse practitio macbook ...

Group 7:

29098: fancy something a bit different? why not try the @drygate #apple #ale? brewed in glesga toon! #glasgow #craftbeer

35679: this #apple is #amazeballs good. cc @umnews @sewardcoop

5. Conclusion

The experiment results show that the proposed method for clustering is effective for the tweets subtopics analysis. From the project, we find out that the SVD operation on sparse matrix is fast. In addition, we use Silhouette value to evaluate results. Then we can automatically adjust the number of clusters.

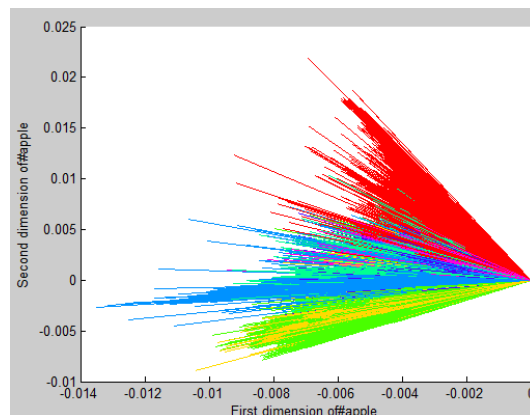
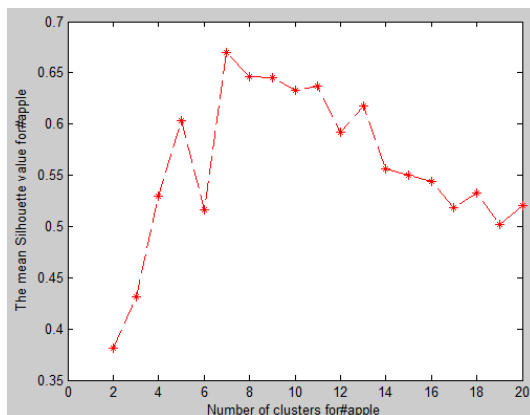
References

- [1] Cha, Meeyoung, et al. "Measuring User Influence in Twitter: The Million Follower Fallacy." ICWSM 10 (2010): 10-17.
- [2] Tsur, Oren, and Ari Rappoport. "What's in a hashtag? Content based prediction of the spread of ideas in microblogging communities." *Proceedings of the fifth ACM international conference on Web search and data mining*. ACM, 2012.
- [3] Twitter Developers. <https://dev.twitter.com/>
- [4] Twitter4J. <http://twitter4j.org/en/index.html>
- [5] Lucene. <http://lucene.apache.org/>
- [6] Dumais, Susan T., et al. "Using latent semantic analysis to improve access to textual information." *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 1988.

Appendix

Clustering results for #apple, #google, #microsoft, #michigan, #cometlanding are listed here. For each hashtag, we show the mean Silhouette value diagram and clustering result with colors in another diagram. For each hashtag, 50 tweets are randomly selected. Each tweet is added to a cluster group.

#apple (7 clusters)



Group 1:

11239: there is no suspense: scotus has already decided to kill obamacare #news #phone #apple #mobile

33920: new southern miss logos feature different eyes #news #phone #apple #mobile

37415: hertz: we actually made \$ million less than we thought (htz) #news #phone #apple #mobile

47422: fortune magazine: members of congress are apple investors . more #apple #iphone #news -

17174: oil prices are falling #news #phone #apple #mobile

64361: here?s how to de-register imessage if you?ve switched platforms with? #mobile #tc #apple #imessage #techcrunch #tech

19162: latin american coffee farmers prosper in a new partnership with chick-fil-a #news #phone #apple #mobile

Group 2:

7462: spain collections . intensive english s - sit tv production unit #itunesu #itunes #iphone #apple #mac

5528: china courses . stage japanese - kolbe catholic college #itunesu #itunes #iphone #apple #mac

18592: france courses . machine learning - stanford #itunesu #itunes #iphone #apple #mac

15353: united states courses . ipad handbook - regina dominican high school #itunesu #itunes #iphone #apple #mac

Group 3:

51586: #apple itunes . . for windows xp and free download: apple itunes? @getsoftw download

28937: united states game paid . dude perfect - dude perfect error #itunes #iphone #apps #apple

62626: germany game free . tetris? blitz - electronic arts #itunes #iphone #apps #apple

32418: germany paid . winmail viewer - enolsoft co. ltd. #mac #itunes #apps #apple

6207: united kingdom game free . the simpsons?: tapped out - electronic arts #itunes #iphone #apps #apple

67647: the game (season). hd premiere now on itunes. [\$.] #apple

Group 4:

26158: veaware of the #masqueattack #iphone #iphad #ios #apple #technology

20229: #icloud #ios - when apple first launched the icloud photo library beta alongside ios . the... - #apple #rumor

19634: imore show : taking a snow leopard moment #ios #apple

50726: apple retail making gift card and uniform changes for holiday shopping season #mac #apple #rumours #ios #osx

28660: #ios #jailbreak now compatible with macs #apple #ios #tech #news

64760: top #ios news: no #iphone plus recall office for iphone #apple watch to be big bucks

24264: roundup: the best third-party keyboards for ios appleinsider #apple

42229: ios vs. #android lollipop: same features yet more distinct than ever #google #apple

Group 5:

67013: #apple #iphone - hybrid id credit card holder case cover with stand for apple iphone plus bl... #deals ebay ca

46812: we fix all smartphones and tablets #iphone #iphone plus #samsung #apple #galaxys #nexus #htc #iphone #screen

56977: #apple #iphone - gold color style hard plastic cover case for iphone plus: au \$. end date... #deals ebay au

56013: #apple #iphone - green . mm super thin matte transparent protective case cover for iphone : ... #deals ebay uk

67130: i want an iphone + so bad it hurts #apple

46990: #ebay #apple iphone - gb - black & slate (verizon) smartphone-bundle

36358: iphone vs. entire galaxy line: will apple win the holidays? #apple #iphone #news

54046: m iphone charger wire has got to be one of the best things i ve brought #apple

Group 6:

10899: new #apple #macbook - item - smile if you love an arts administration. sleeves for macbooks

46029: yahoo!auction:apple a a a md md md ssd usb . ? #???? #apple #???? #????? #sougofollow

53343: trident case aegis case for apple ipad mini black #ipad #apple #case

44368: new #apple #macbook - item - smile if you love a gerontological nurse practitio macbook ...

Group 7:

47390: almost done with my #savory #apple #pie ! #savorypie #applepie #bacon #thyme #fall #pnw #seattle?

24513: for a low low price on new #apple #external #harddrive fast delivery! buy now! apple external harddrive

25973: tech logos before they became iconic... #apple #nokia #ibm #xerox #motorola #microsoft

29098: fancy something a bit different? why not try the @drygate #apple #ale? brewed in glesga toon! #glasgow #craftbeer

3643: fitbit products remain on the shelves at apple stores for now #apple

16847: #apple #ios #app #in video-sharing app ocho puts your new iphone?s supersized screen to wor... #iphone #ipad #mac

33216: photo: apple voodoo in #blankspace. this one has an extra green one. @taylorswift #taylorswift #apple

52226: belay cords #apple #ausgefallen #design

28527: an apple a day. #apple #nomnom

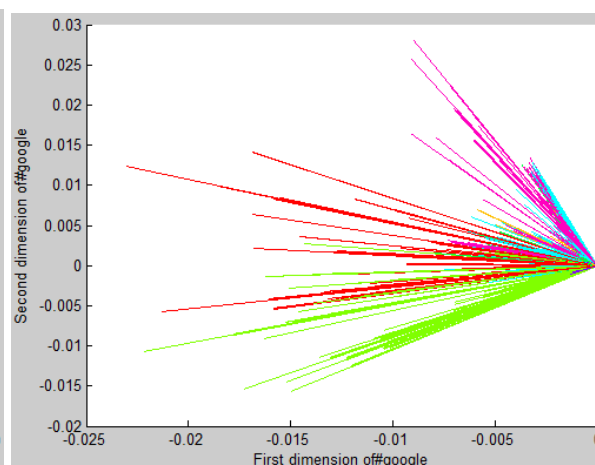
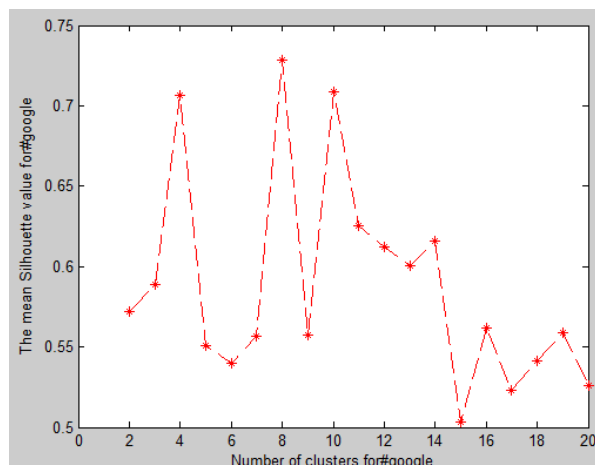
55292: #design #supermoto #apple how to #build #auto lawn sprinklers.: how to design install and build your own aut...

47360: here it is! my yr apple anniversary. it s been a crazy ! #appleversary #apple #timcook # yrs?

35679: this #apple is #amazeballs good. cc @umnews @sewardcoop

33938: #latenightssnack #bestyoghurt you ll ever have #noosa #yogurt with #honeycrisp #apple and #cacao?

#google (8 clusters)



Group 1:

37398: @heyrrayy the first project came up! technology application.. #google

43705: @budakkulimbb the first application came up! flashlight application.. #google

Group 2:

47526: #google matt damon to reprise role in th bourne movie - people magazine

28656: #website #stresstest #google #ranking how does your site rank

19295: just installed google messaging and google keyboard. so much material goodness #google

40867: #nexusplayer by #google announced!

18488: link to the new #google maps . . #apk (it s really nice too...)

40178: #google benefits because more people have to rely on #adwords to get the same kind of #traffic

2707: only disney is magical enough to get #google and #apple to make a deal #disneymovies

52092: the recent google drive updates teachers should know about #google drive

15528: kobe bryant sets record for missed shots as lakers fall to grizzlies - - los angeles times #google #sport

38446: . #google co-founder #sergeybrin is third. according to the latest #forbes

114: #google sniper get your website on page the number one selling page rank submission software

37517: #google has a #hangout zone! find here

32470: vmware shows up at aws event hawking hybrid cloud management - crn australia #google #news #tech

Group 3:

36141: @mistergreenwich will you support me by downloading my program? #google

45627: @kyle saechao download and support me please. :) #google

Group 4:

26543: ?@socialmedia day: will shopping ads be the knowledge graph of paid search? via @larrykim #google

332: #google to take over #moffett airfield and invest \$ m+ to refurb hanger and other facilities

43224: #google wpi explores using robots to treat ebola - worcester telegram

23662: who still need #google #inbox invitations ???

10545: despite ek s statement #spotify goes down but #google goes up thanks to the mini revolution - youtube music key

490: facebook is taking over from #google as gatekeeper of the internet via @betanews #social #facebook #censorship

6307: search with #google & enjoy. #gonawazgo #babymaryamgotoschool #shameonlumsvc #pakistanlovesasadumar #aikbaatsuno

15146: got my #google #inbox invite and i m not a fan of how my emails are grouped by default. it s going to take some work.

12021: google s new youtube music service is here but invite-only for now via @mashable can.not.wait #google #streaming

11294: #google s new #youtube #music service is here but invite-only for now via @mashable

Group 5:

16566: now trending on sd times: #google releases tools for developing #android apps in c/c++

20561: #android #google #news #android lollipopotaupdate android . lollipop ota update for nexus wifi (&?

40073: @iambulb the new arge mobile app appeared. #google

24955: #google releases standalone #messenger app for #android.

46610: @jrod the new android mobile app appeared. #google

604: nexus reported to have limited supply available to phone carriers ? #google #tmobile #sad

10633: where is my god damn lollipop update #android ? #nexus #waiting #google

Group 6:

19494: @intrasource we provide development & maintenance services.visit for more details #software #google #outsourcing

24936: buaireland: google wants us privacy act powers extended to eu citizens #google #snowden without snowden how wou...

16155: mercedes-benz recalls c-class vehicles over steering issue - auto world news #google #news #tech

6786: first world war in #google street view #rememberthem #war

Group 7:

19982: congress returns for lame-duck session #lemonde #google #reuters

51827: iran: another acid attack threat in ahwaz #google #iran #world

43781: gen. martin dempsey discusses iran's nuclear ambitions #google #iran

41017: iran: narges mohammadi lawyer and activist has been summoned for interrogation #euronews #iran #google

5704: kin of dead us soldiers sue banks over #iran dealings #google #usa #irantalks

2315: iran: unexpected transfer of political prison afshin baymani to ghohardasht prison #iran #un #google

43274: gen. martin dempsey discusses iran's nuclear ambitions #newsfeed #google #headline

24395: obama seeks review of #syria strategy sees assad removal as necessary: cnn #iran #google #world

26747: special report: the fighters of #iraq who answer to iran #iran #woman #google

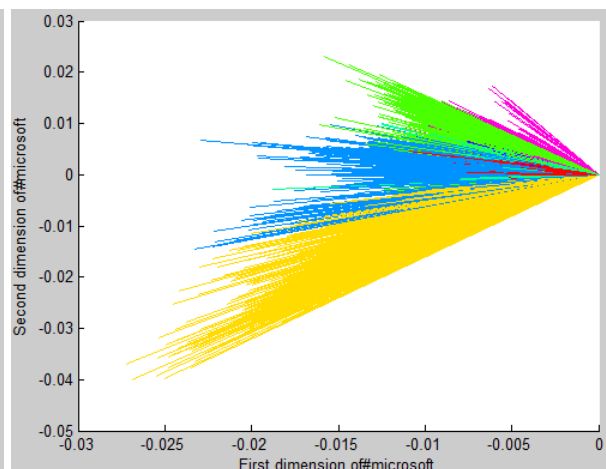
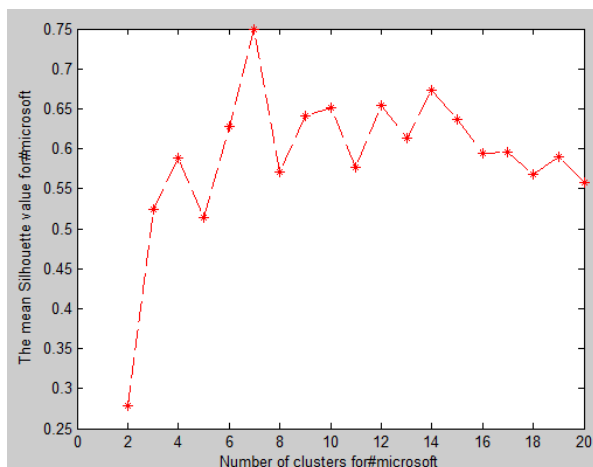
23961: tehran: youth clashing with suppressive security force #health #breakingnew #google

Group 8:

27424: #google apps for education #gafesummit #edtechteam #gafesummit in istanbul hisar schools

41207: @rsvine are flashlight app tries to please ? #google

#microsoft (7 clusters)



Group 1:

18586: although one of my fave things being a convert to the #microsoft ecosystem is the value i see myself trying to overspend anyway ha

11056: the best microsoft outlook... #microsoft #outlook

10368: #mtnmsftcloud: simon ouattara ceo of #microsoft west central africa reminds that #africa is the continent for future

17792: #education #microsoft microsoft innovative educator expert very honoured to have received this! thank you microsoft

17017: #microsoft #senior ux designer - bing ux

11409: new post: #contests #featured #halo #halo #microsoft gr giveaway ? halo: the master chief collection swag!

11081: all the new software & previews #microsoft launched is awesome now i just #needmoretime to play with all of this. #vs #dotnet

16869: the new #microsoft....

8712: new post: #band #gadgets #microsoft microsoft band division?s head has custom band

18464: #toys ...i might be in this market...it s such a cool device #microsoftband #microsoft #electronics?

16308: amd launches the radeon r x with gb of graphics memory #microsoft #amd

12171: if only #microsoft were smart enough to release it on pc. oh well.

8320: #microsoft says xbox one sales have ?skyrocketed? in recent weeks

18720: bought them from #ebay in the end! #minecraft - #microsoft sot it out! girls love minecraft - know your customers!

Group 2:

13501: ? lumia collection ? #lumia #microsoft #nokia #windowsphone

4358: cool ? news. #microsoft #lumia is official now. introduction video. price and availability

12325: new update wp . battery ? #microsoft #lumia #nokialumia #morelumia

7837: #microsoft #lumia hitting the market today...gonna be at affordadble price..

4805: #microsoft relinquishes the #nokia name for its low-cost lumia #gaming

5808: anyone know what i am looking at? #nokia #microsoft #microbiology

17368: #microsoft to launch first non-nokia smartphone on nov - times of india

Group 3:

19145: #microsoft offers prorated #office refunds to paid subscribers after making mobile editing free

18893: how #microsoft can afford to give away #office on ios and android via @feedly

18254: microsoft sets office free on iphone ipad with android preview and windows ... - gigaom #microsoft

1159: #enterprise #storage #tc #microsoft #office microsoft?s ceo breaks down the new soul of?

9370: #crowdfund #investors to help you develop your #mobile #app #apple #microsoft #android

Group 4:

12109: magic costume and garage of mystery make for halloween hardware maddness #microsoft

16381: microsoft has a new xbox bundle with a free copy of peggle for \$ #microsoft

12779: #microsoft doesn t plan on making the same mistake twice: the company has made a public pledge that all of...

15089: hands on: #microsoft #sway tells #your #story without the hassle

14870: #microsoft is now second most valuable company behind #apple

7656: should microsoft build hardware? #microsoft #windows #wp #satyanadella #surface #azure #itsupport #tablet #office

16567: our customers are always asking us for device recommendations...check out the newest pc! #microsoft #pc #mobile

682: here?s what #microsoft ceo #satyanadella thinks #apple and #google do?best via @techcrunch

10617: fancy a stunning new #microsoft #surface pro for under ? #deals #computers #auction #offers

Group 5:

4171: microsoft ends retail sales of #windows #microsoft

12276: ibm just found a bug in every version of windows since - #ibm #microsoft #security

19777: #johnlewis and #microsoft build monty s magical toy machine. sounds charming cap duly doffed

12144: #microsoft buys #israel s #aorato; specialises in pointing out security flaws in products. #bddfail #victims

Group 6:

15567: itunes:gaming cyclone xbox t&g crimson viper gaming cyclone xbox t&g #xbox #microsoft #sougofollow #??????

9570: xbox is trying to catch up to ps sales. #ps #xbox #videogames #microsoft #sony #console

2408: #skype for business to replace #microsoft #lync: via @verge

12369: #microsoft s #screamride goes off the rails in march - #xboxone #xbox #xboxlive

Group 7:

2537: @si fantasy the jacksonville family robinsons via @jaguarscom #microsoft

6388: visual studio community - thank you #microsoft

8317: if you wonder whether #linux is a good operating system note that #microsoft which sells #windows runs their #bing search engine on linux

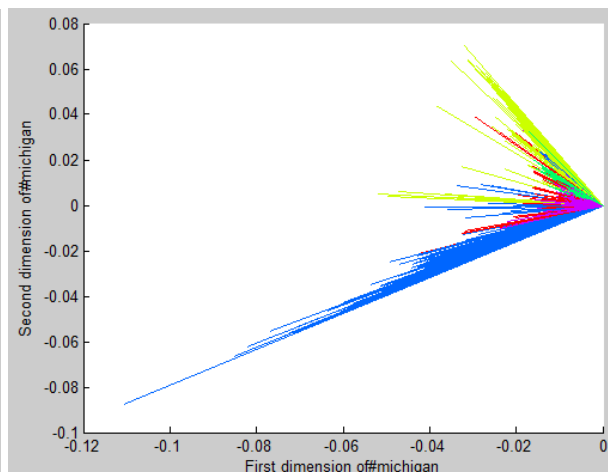
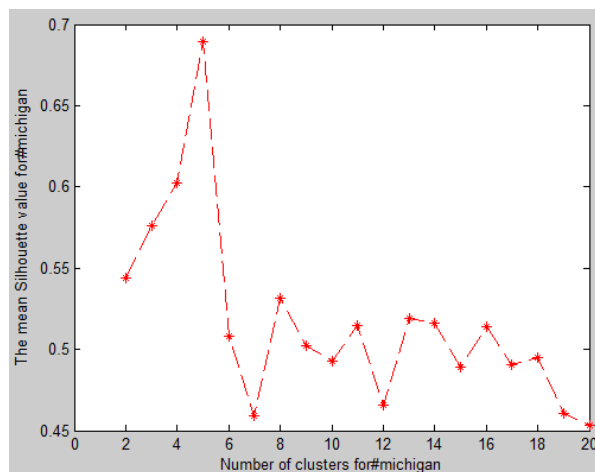
9711: @ers @kaepernick @aldonsmith rudolph the redzone viking via @jaguarscom #microsoft

10591: #microsoft makes #.net #opensource available on #mac and #linux.

6636: via @natfriedman: congratulations to #microsoft on open sourcing .net and releasing a free version of visual studio! #opensource #ftw

10268: #microsoft: .net wird open source - the right step in the eyes of an ex .net developer. maybe too late...

#michigan (5 clusters)



Group 1:

9294: looking for something to do in #riverside #michigan? check out this great new mobile app!

4200: what s going on in #ferrysburg #michigan? check out this great new app for news and weather!

Group 2:

6076: day #happydays downtown #detroit #michigan #happydayschallenge #blackandwhite?

101: new music she like in my bio now!!! #music #michigan #det #detroit #teamemu #theredkarpet #la #losangeles #los...

3619: detroit (wwj) ? a film shot entirely on location in ann arbor the campus of eastern michigan? #detroit #michigan

146: the bidding start time for this property will begin on december th! #detroit #michigan

346: are you looking to earn a college football scholarship? #michigan #mhsaa #highschoolfootball #football #detroit

401: signs you're from? #michigan via @thoughtcatalog #usice #puremichigan #detroit #michigan happy #monday!

2603: you lookin at me?? #rescue #adopt #gsd #wgsd #dog #fall #detroit #michigan #walk #prince #love #puppy credits go...

5742: me and my cousins at mgm casino #mgm #michigan #naseemjalal casino @ mgm grand detroit

5588: detroit barber sebastian jackson gets national tv attention tonight deadline detroit an? #detroit #michigan

Group 3:

8034: #michigan/#northwestern game putting me to sleep. - with seconds to go in the half. #seriously

7624: a war of attrition is an understatement for this #michigan game. better description-- fugly. #umvnmw #umvsnw

8190: so confusing #michigan @bobwojnowski: no idea why smith is getting key carries over drake johnson. no idea. michigan football -- no idea.?

Group 4:

6541: accelerate #michigan #innovation #competition #announces #winners [xconomy vc/deals]

2932: bargain! reault #flint #michigan tell a friend.

5937: #detroit #michigan nov : temperature jãf light snow wind w km/h humidity % ..

Group 5:

1760: @iamjericho can t wait to see y j in rock and roll form #y j #fozzy #michigan

5081: he s been doing this for years at #michigan now he s doing it on a bigger stage for everyone to see @treyburke #nba

1078: if you or a loved one have been injured or killed due to the negligence of another person in #michigan #call #thurswelllaw at - - .

2235: former marine explosive-detecting dog from west #michigan celebrated at america?s parade in new york

2274: reverend pinkney: the civil rights hero you haven t met ? @webstergtarpley #mlk #michigan # u

2700: this is not real #michigan

5424: love my mom!! #shespretty #family #michigan

8323: #ncaafb adding #iowastate - - (pm pst) #michigan + - (pm pst)

8240: stoops #ou to #michigan? stay tuned...

1822: #michigan dnr adds more aquatic species to prohibited list including killer shrimp.

5867: we are all #drivingamerica #thehenryford #dearborn #michigan

4071: #grandrapids #marijuana decriminalization goes to state appeals court today. #mi #michigan

792: isenbl?st (u.s.) a fire of knowledge and power (promo video) #isenblast #blackmetal #unitedstates #michigan...

3778: not like we needed a list to prove it but here it is! #michigan

6195: i m in love with our great state of #michigan so if you are as well # join in and help me spread the?

8815: @desmondhoward locker room tour today #goblue #michigan #heismanvote

5409: wittle baby avagrace. # x #michigan #igmichigan #michigangram #familyphoto maggiensch

3891: help give the gift of hope to michigan patients waiting for an organ transplant... #donatelifemi #michigan

1564: final: #michigan tops wayne state - in exhibition.

7259: police tracking dog finds missing -year-old girl in #michigan

2040: just accepted a full time assistant coaching position at davenport university coaching all jumps and hurdles! #excited #polevault #michigan

4368: really impressed with #rightbrainbrewery from #traversecity #michigan. enjoying the #spinaltapper?

2332: little known pdf of african studies center at michigan state university #african #studies #center #michigan #state

3457: urgent! stop political lynching of #revpinkney #michigan @redactedtonight @abbymartin @edschultzmsnbc

3756: it's cold here in #michigan! wear your @lhfashion under a #sweater and #staycool when you turn up the heat #menswear

2358: job: team leader / registered nurse #hiring #nursing #healthcare #resume #employment #michigan #tweetmyjobs #hr #job

8442: if #oklahoma head coach #bobstoops can get out of dodge and get a job like #michigan or #florida it would be a good career move. #gonestale

6004: #michigan demonstrates why #elections matter #tcot #teaparty

2154: kid kicking cancer empowers children in fight for their lives: kids kicking cancer helps kids fight ... #michigan

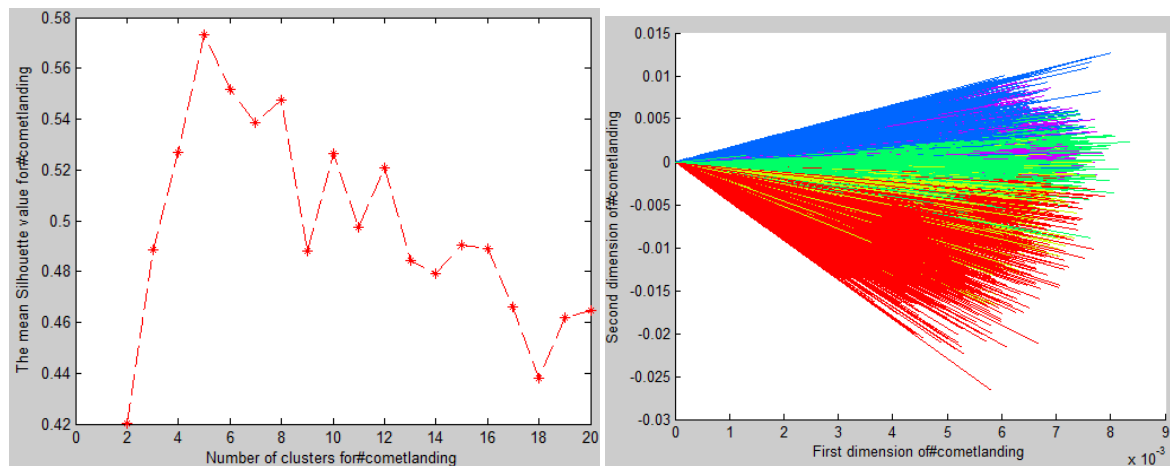
1958: does your house need a homework central zone? look what this #michigan family created

4751: hello my friend. chelada yum #chelada #michigan @ the gameroom at the peruvian palace

8217: eek that is close. they are marking him short for a th and . what does #michigan do?

1292: so it begins! #winter #michigan #cold

#cometlanding (5 clusters)



Group 1:

49150: in coming #cometlanding

59849: so many people watching this historical moment that all the streams keep freezing and crashing #fingerscrossed #cometlanding

69570: with vision + cooperation we can even do things like this! #rosetta #cometlanding touch down c : gmt **live**

68457: #cometlanding is a historic occasion. see to it your kids make history check out the #homework helpers series

33727: really amusing to see how the scientists #cometlanding weren't ready for the microphones - endearing

22617: #rosettamission #cometlanding .now i am watching on #bbc scientists first moment .

88190: group greystones library drawing # p #actioncomet #sciwk . lunch then group all go here. #cometlanding

1573: the major event #cometlanding

1140: message from william shatner - aye aye captain! #cometlanding

33507: the biggest problem with success is that it looks easy especially for those of us who are just watching. #cometlanding

49989: oh my... almost half a million people watching the @esa rosetta landing webcast at #cometlanding @philae

Group 2:

45631: #cometlanding yay hugs and applause handshakes. must be good news. congrats

5465: for the day that s in it: #cometlanding #adorbs

7274: still waiting for the hugs applause and tears of joy suddenly turned to muffled gasps of horror plot twist #cometlanding #rosetta

8973: cool #cometlanding and good to see a science #bbcnews headline

40126: the cheering was a good sign! #cometlanding

31287: an exciting day for european space agency !! #cometlanding

Group 3:

77945: i wonder if they can turn the camera around on @philae to take a pic of #rosetta ? #cometlanding

81641: ?icymi @nasa live @esa rosetta s @philae set for st ever probe #cometlanding: #toshare

2103: @esa rosetta mission readies for @philae #cometlanding ?we've been doing science on side.?

75364: @tomdearie @pompeycal are you following @philae and @esa rosetta ? brilliant. #cometlanding

30322: @ukip can you see what european integration can achieve? your isolationism would have given us only ignorance. #rosetta #cometlanding

4109: ?@team maths: yet more #cometlanding maths! worth every euro/cent! # mathsdays

37185: congratulations to the @esa @esaoperations @esa rosetta! #cometlanding

24427: @esa well done!! the future is now! #cometlanding

37607: #cometlanding succes!

Group 4:

83653: where have people got the idea that # p is billion miles away?yes it s far but only million miles. #cometlanding

20378: sure the #cometlanding is impressive in a conventional sense but talking walking pants grappling with morality?

40568: from our rock to a hard place #cometlanding

8764: kim kardashian has totally stolen #cometlanding s thunder. #putsomeclotheson

88021: cheap as chips! mt @scienceogram: how much to land on a comet? @esa rosetta #cometlanding

85906: this dude just helped land a spaceship on a comet. what an absolute legend. #cometlanding

49448: ??????????#cometlanding #breaktheinternet warren clarke #datass

15455: why is the rosetta comet landing so exciting? network front | the guardian #cometlanding is trending and we hav...

90246: #cometlanding is the best reality-show ever.

79421: is that stable as in intensive care but not quite dead or stable as in we tested this software for weeks solid ? #cometlanding

28681: #cometlanding a great human achievement. now let s land a probe on poverty in the eu. whitey s on the moon.....

59757: #cometlanding today landing on a comet. tomorrow mastering the force!

43774: pics or it didn t happen! #cometlanding

91129: well they ve put a dishwasher on a comet so... - my mum on the #cometlanding.

18139: shhhh. its ok tell your inner nerd to stop running in circles and hyperventilating. its super cool so stop squealing #cometlanding

66925: randall munroe is the man. #cometlanding

43126: twitter is breaking.... help #cometlanding

21623: wonder why #wecanlandonacometbutwecant isnt trending yet? #cometlanding

55769: @randamma first for us humans/ st space craft to land on a comet!!! #cometlanding / though you should know/.

25696: amazing! for the first time in human history a spacecraft has landed on a comet. congrats to the @esa! #cometlanding

Group 5:

14060: exciting!!! touchdown confirmed for philae at : cet ? rosetta - esa s comet chaser ##cometlanding #cometl [...]

85043: fight damn it! fight! @philae #cometlanding

56: videos of the week: this week? @philae naturally! & stay tuned on nov th! #cometlanding

2179: #cometlanding philae lander is packing for the trip. only a few hours until launch!