The Effect of Keywords Used on Content Attraction in Complex Networks

Yassine El Moudene*
Jaafar Idrais
IMI Laboratory, Faculty of Sciences of Agadir
Ibn Zohr University, Agadir, Morocco

Abderrahim Sabour
Department of Computer Science
High School of Technology, Ibn Zohr University, Agadir, Morocco
yassine.elmoudene@gmail.com

To understand complex networks like online social networks (OSNs), one must explore them. This gives a partial view of the real object, which is generally assumed to be representative of the whole, like an iceberg. OSNs are growing rapidly, not only in the number of users they attract, but also in the mass of data produced in a short period of time. Once created, the data attracts the attention of other users, who interact with it. The purpose of this study is to evaluate the impact of keywords used on content attraction, through an analysis of the attraction's different modes as well as the nature of the interactions of the different communities. The study concerns a crucial topic: the environment and climate change. This paper looks at the keywords used on Twitter and their impact on the interactions of different users for this complex network.

Keywords: online social network; big data; PCA; content attraction; keyword effect; complex system; data extraction; data analysis; knowledge extraction

1. Introduction

Nowadays nobody can deny the impact of online social networks (OSNs) [1] on the world, especially big corporations such as Facebook, Twitter, LinkedIn and Instagram. The OSNs [2] are a growing phenomenon due to the variety of characteristics of each one. Facebook, for instance, is characterized by a reciprocity in relations between users: the relationship in this case is unsupervised. On the other hand, on Twitter that is not the case: members may not necessarily have a reciprocal relationship with other members. In this case, the relationship is directed or not directed. In our study on data analysis, and for a variety of reasons, we focus on Twitter, which is an online social network [3] that allows users to send and read

https://doi.org/10.25088/ComplexSystems.29.3.689

140-character short messages called tweets. Twitter is accessible to unregistered users who can read and interact with most tweets, unlike Facebook, where users can control the privacy of their profiles. The massive information provided by Twitter, such as tweets, user profile information, the number of subscribers (followers) and the number of subscriptions (accounts followed) in the network, plays an important role in data analysis. Previous studies were interested in the effect of network structure [4]. In this paper, we analyze the impact of keywords used for highlighting tweets and attracting the attention of users via an analysis of the interactivity of different modes.

2. Methodology

In this section we describe the methodology and data object used. Principal components analysis (PCA) [5] is used to study and analyze the Twitter community data from the definition of variables [6] to the analysis of results by extracting factorial axes [7]. A graphical representation is given that reflects the distribution of individuals as well as variables, using a verification and validation processing procedure based on source data.

2.1 Dataset

The traditional relational database management system (RDBMS) can process data when it is well structured. With the increasing amounts of unstructured data in various sources (especially on OSNs [8]) considered as big data, RDBMS cannot be used for processing unstructured data, hence the need for NoSQL databases. There are different types of Twitter data, such as a user's data: their own personal information, accounts followed, followers, tweets and their reactions—likes, replies to other tweets, retweets.

2.2 Data Lake versus Data Warehouse

Data lakes and data warehouses both are widely used to store large amounts of data. A data lake is a vast set of raw data whose purpose is not yet defined. A data warehouse is a repository of structured and filtered data that has already been processed for a specific purpose. Both types of data storage are often confused, but they are much more different than they look. In fact, the only real similarity between them is their high-level objective of storing data. Data warehouses have been in the field for almost 30 years. Recently however, data lakes have gained popularity. In fact, we should keep in mind that these two types of systems have important differences and are not used in the same way. Table 1 shows some of those differences.

	Data Lake	Data Warehouse
data structure	raw	processed
data purpose	not yet determined	currently in use
users	data scientists	business professionals
accessibility	highly accessible and quick to update	more complicated and costly to make changes

Table 1. Comparison between data lake and data warehouse.

2.3 Extraction Process

In our study, we choose to target the interest of the Twitter community in the topic of the environment and renewable energies [9], in order to do a comparative study based on three axes using 45 keywords divided into three categories, as explained in Table 2.

Category	Keywords
English	renewable energy, electric vehicle, electric cars, CO2, climate change, smart home, green energy, global warming, air pollution, GCAS2018, green France, climate action, green Morocco, greenhouse gases, global climate action summit, green_morocco, Make Our Planet Great Again, solar energy
French	éoliennes, maroc_vert, solaires, energie renouvelable, hydroélectricité, changement climatique, rechauffement, climatique, biomasse, photovoltaique, fossile, nucléaire
Arabic	الطاقات المتجددة،الاحتباس الحراري،الأحفور،تغير المناخ، كوب ٢٢ قمة باريس للمناخ،اتفاق باريس للمناخ،محطة نور للطاقة الشمسية النفايات المستوردة من ايطاليا،الطاقة الشمسية ،الطاقة الخضراء الطاقة الريحية ،قمة المغرب الأخضر، زيرو ميكا ،بونظيف ،كوب ٢١

Table 2. Keywords categories.

First of all, the keywords constituting the extraction [10] queue were chosen, then we used the Twitter search page to launch a search of different tweets containing the keyword in question (Figure 1). For each tweet we collected [11] its specific information, namely: owner identity, number of likes, number of replies, number of retweets, number of followers and number of accounts being followed.

2.4 Dataset Statistics

At the end of the collection process, we were able to retrieve 107 314 tweets, from the first launched on Wednesday, March 28, 2007, at 18:28:47 GMT by Mr. Tozster (joined in December 2006), until the last tweet on Wednesday, September 26, 2018, at 08:53:07 GMT by a certain Baptist Romeuf (joined in January 2012). Table 3 illustrates the different dataset records.

Table	Records	Interactions	Number
tweets	107 314	likes	994 744
followers	494 693	retweets	1 106 224
followed	385 522	replies	418 089
users	77 830		

Table 3. Dataset records statistics.

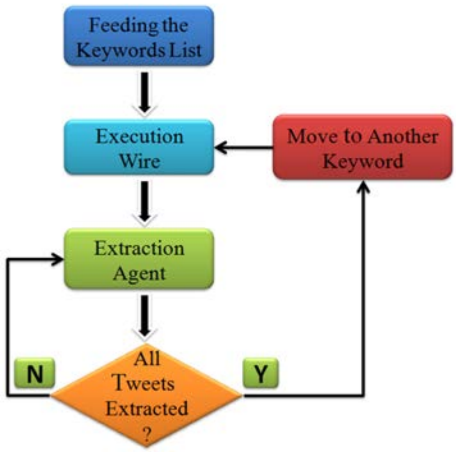

Figure 1. Semi-supervised data acquisition process.

3. Preliminary Analysis

Before analyzing our data, we first carried out a preliminary analysis of the collected data. Since our study is based on the analysis of keywords in relation to the interaction of different users, we first examined the interactivity rate and its evolution, then we tried to visualize the collected data distribution at the temporal scale.

3.1 Different Reactivity Modes Rate

On Twitter, a user has three possible choices to respond to a tweet: “like,” “reply” and “retweet.” According to Table 4, 89% of tweets extracted were liked and retweeted, while 61% got a reply, so “like” and “retweet” are the most used interaction modes. Focusing on the evolution of the interaction modes, Figure 2 shows an exponential evolution of retweets and likes, while we notice a weak growth in the replies (responses) mode.

Interaction	Number	Reactivity Rate
liked tweets	96 242	89.60%
retweeted tweets	96 363	89.79%
answered tweets	66 292	61.77%

Table 4. Different reactivity modes rate.

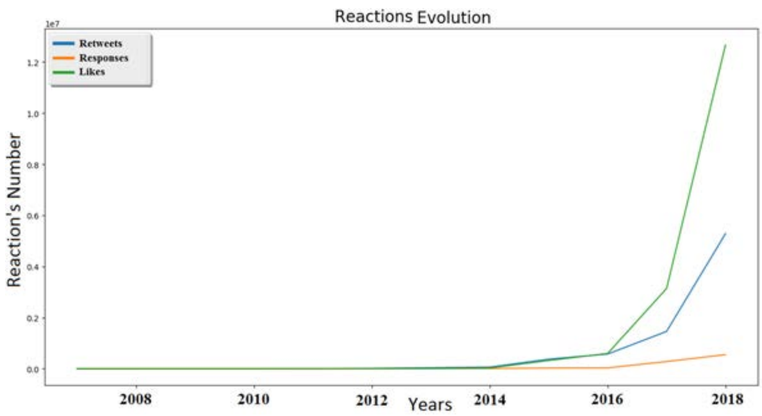

Figure 2. Evolution of reactivity modes.

3.2 The Evolution of Interest in the Studied Domain

Figure 3 illustrates the evolution of the number of tweets per year. We note an exponential interest from the year 2014, which means that the international community has started to take an interest in the environment and renewable energies [12] topics since that year.

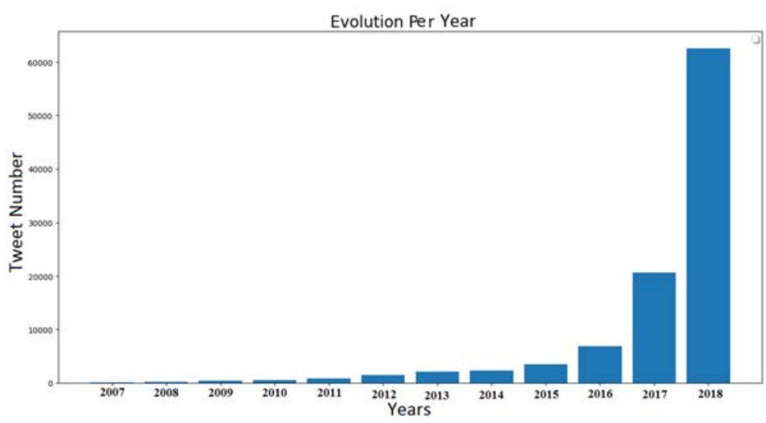

Figure 3. Interest evolution over years.

The interest found in the previous section is validated by Figure 4, which zooms in on a month window for each year. There are three categories:

- From 2007 to 2014: Low interest in the environment and renewable energies [13] subject throughout the year.
- From 2015 to 2017: The international community begins to focus on climate change and renewable energy; this interest starts in late May and early June, spreading over the remaining six months of the year. This change is probably due to the preparation for World Environment Day, which is June 5.
- 2018: Where we notice that the environment topic begins to grow via the quantity of data, and that interest had already begun in January and February.

From Figure 5, we notice a specific interest in the field of study during the summer season, a period of travel and vacation.

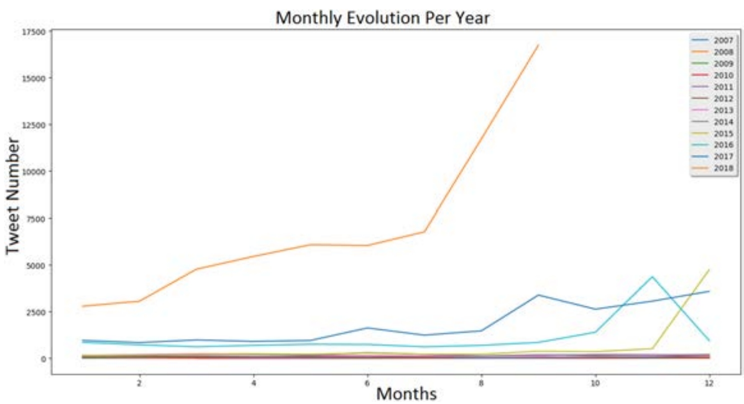

Figure 4. Interest evolution per month over the years.

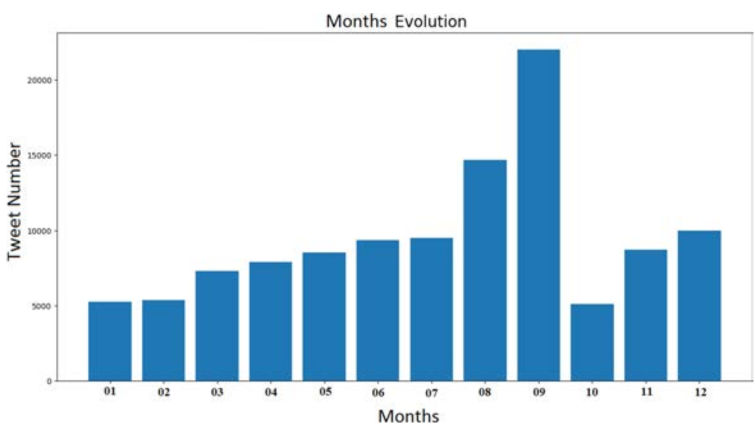

Figure 5. Interest evolution over the months from 2007 to 2018.

3.3 Moroccan and French Interest in the Environment

France and Morocco had a quick surge in interest over the two years 2015–2016. A deep research analysis of extracted data relating to these two years shows that this growth was due to France's preparations for COP21 (November 30 to December 12, 2015) and Morocco's for COP22 (November 7 to November 18, 2016).

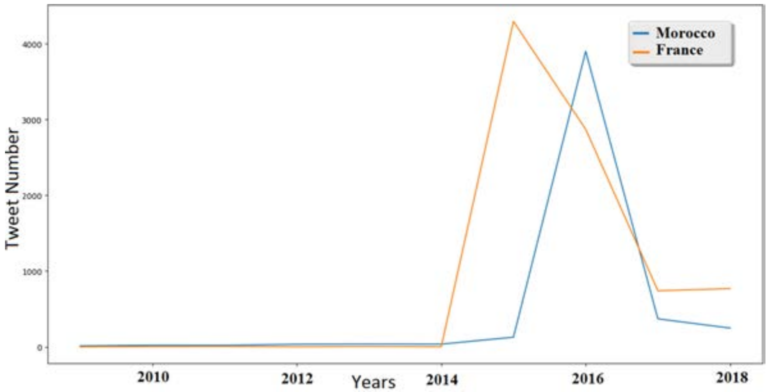

Figure 6. Comparison between France and Morocco.

4. Keywords Analysis

The objective of our study is to evaluate the impact of the keywords used on the content attraction, through an analysis of the attraction's different modes as well as the nature of the interactions between different communities using keywords reserved for the environment and renewable energies field. To do this we chose the PCA [14] as a model. At first we study the preliminary relationship [15] between the different keywords, after that we analyze these relations, and finally visualize a graph in order to validate the results.

4.1 Keywords Edge Weights Analysis

Let m_1 and m_2 be two keywords; tweets_{m1} and tweets_{m2} are, respectively, the extracted tweets using m_1 and m_2. The relation $R(m_1, m_2)$ is the number of tweets in which those two keywords appear:

$$R(m_1, m_2) = \text{card}\left(\text{tweets}_{m1} \bigcap \text{tweets}_{m2}\right).$$

Figure 7 shows the distribution of relation R measured between different keywords. We notice that the maximum relation $R = 244$ which, while fetching data, is between the two keywords (renewable energy, fossile), that cross 244 times.

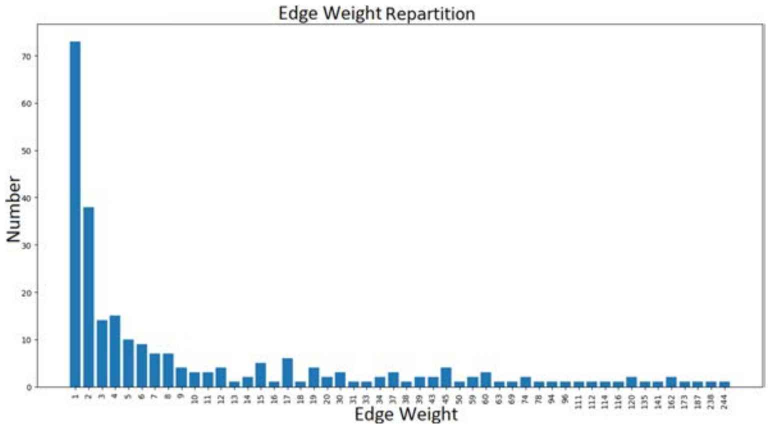

Figure 7. Weights between keywords.

4.2 Keywords Relationship Visualization Using ForceAtlas2

ForceAtlas2 [16] is a spatialization algorithm implemented in the Gephi software. It allows for a force-directed layout. The nodes repel each other, like charged particles; while the edges attract their nodes, like springs. These forces create a movement that converges to a balanced state. Figure 8 represents a weighted graph, the nodes represent the keywords m, and the edges are the measured relation R between m_1 and m_2. Notice the size difference between some nodes/edges, and also the dispatching of the nodes. So we measure modularity, using the Louvain method implemented in Gephi, in order to extract the network structure. Modularity is designed to measure the strength of division of a network into modules (also called communities). Figure 9 shows that six communities are detected from our graph with a low modularity (0.233); Table 5 illustrates all keywords that build each community. Communities 1, 2 and 5 are the densest communities,

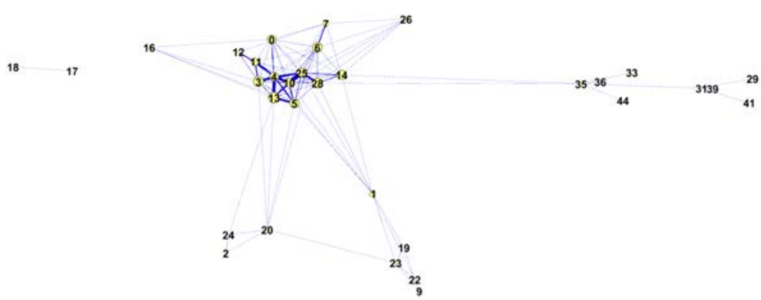

Figure 8. Keywords relationship visualization using ForceAtlas2 algorithm.

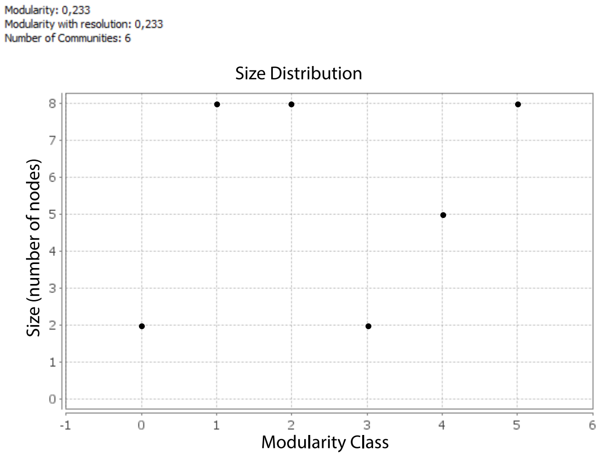

Figure 9. Keywords graph modularity using Louvain method.

with eight nodes for each one: the first is for English-speaking keywords, the second for Arabic ones and finally the French ones. Those communities represent the most keywords used to attract different users; for each community we find a specific common meaning. It should be noted that the two communities 0 and 3 are composed of two keywords that have the same meaning. To better see the measured relationship R, we will apply PCA.

Community	Keywords
0	green Morocco, green_morocco
1	air pollution, climate action, climate change CO2, GCAS2018, global warming Global Climate Action Summit, greenhouse gases
2	اتفاق باريس للمناخ، قمة باريس للمناخ المغرب الاخضر، تغير المناخ، محطة نور للطاقة الشمسية الاحتباس الحراري، الطاقة الخضراء، الطاقة الشمسية
3	electric vehicle, electric cars
4	fossile, green energy, renewable energy, solar energy
5	biomasse, changement climatique, eolienne Make Our Planet Great Again, photovoltaique hydroélectricité, nuclear, rechauffement climatique

Table 5. Keywords communities.

4.3 Keywords Analysis Using Principal Components Analysis

PCA [17] is part of the group of multidimensional descriptive methods called factorial methods. The data is the measurements made on n

units $\{u_1, u_2, \ldots, u_i, \ldots u_n\}$. The v quantitative variables that represent these measures are $\{v_1, v_2, \ldots v_j, \ldots v_p\}$.

4.3.1 Definition of Variables Studied

Our study focuses on different keywords that were the starting point of the information-gathering process. These keywords concern the subject of the environment and renewable energies. Let m be a keyword, then tweets$_m$ represent tweets extracted using m, and U_m represents the different users who made tweets$_m$. Variables are the characteristics of tweets, namely:

- T_m: number of tweets$_m$.
- U_m: number of different users who produced tweets$_m$.
- FR_m: number of U_m followers.
- FW_m: number of accounts followed by U_m.
- L_m: number of likes for tweets$_m$.
- RS_m: number of replies for tweets$_m$.
- RW_m: number of retweets for tweets$_m$.

So for each keyword m we associate the characteristic matrix [15] X composed of $v = 7$ quantitative variables with:

$$X(m) = \left[T_m, FW_m, FR_m, L_m, RS_m, RW_m, U_m\right].$$

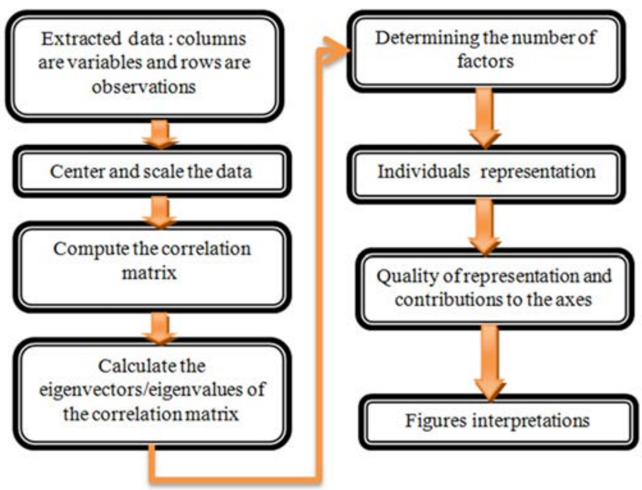

Figure 10. PCA steps.

4.3.2 Data Normalization

We try to extract the information contained in our dataset. To do this, we have to visualize it via a graphical representation and interpret it. Those representations are made in reduced dimension [18]: the initial space of dimension v will be summarized in dimension k. The number of selected factors k will be between 1 and v. To do this, we will create data independent of the unit, and choose the variables having even average and the same dispersion, using the reduced centered matrix denoted H:

$$H_{ij} = \frac{X_{ij} - \acute{X}_j}{\delta_j} \tag{1}$$

$$X_j = \frac{1}{n}\sum_{i=1}^{n} X_{ij} \tag{2}$$

$$\delta_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_{ij} - \acute{X}_j\right)^2}\,. \tag{3}$$

4.3.3 Correlation Matrix

After data normalization, we generate the correlation matrix between those variables. Table 6 shows the generated matrix.

	T_m	FW_m	FR_m	L_m	RS_m	RW_m	U_m
T_m	1.000	0.111	−0.114	0.825	0.810	0.863	−0.084
FW_m	0.111	1.000	−0.086	0.113	0.111	0.113	0.420
FR_m	−0.114	−0.086	1.000	−0.097	−0.086	−0.101	0.039
L_m	0.825	0.113	−0.097	1.000	0.963	0.990	−0.133
RS_m	0.810	0.111	−0.086	0.963	1.000	0.958	−0.111
RW_m	0.863	0.113	−0.101	0.990	0.958	1.000	−0.125
U_m	−0.084	0.420	0.039	−0.133	−0.111	−0.125	1.000

Table 6. Correlation matrix.

4.3.4 Factors Selection

Once done, we continue to calculate eigenvalues associated with the factorial axes. Table 7 shows the generated matrix.

Variables	Eigenvalues	Inertia	Cumulative Inertia
T_m	3.757	0.537	0.537
FW_m	1.423	0.203	0.740
FR_m	0.993	0.142	0.882
L_m	0.539	0.077	0.959
RS_m	0.234	0.033	0.992
RW_m	0.047	0.007	0.999
U_m	0.007	0.001	1.000

Table 7. Eigenvalues.

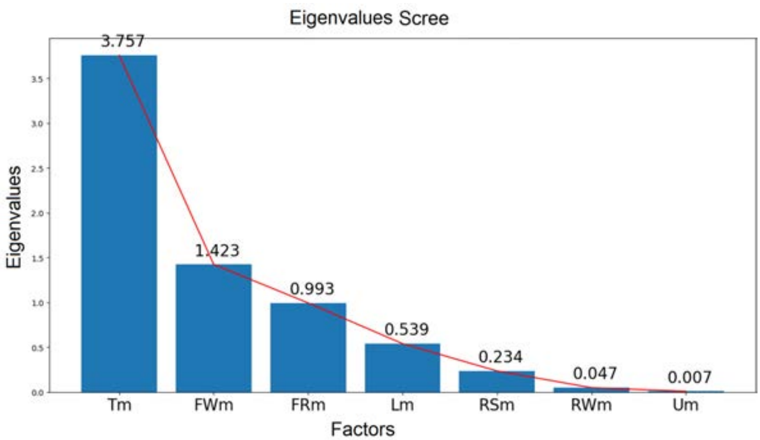

Figure 11. Eigenvalue scree.

The total variation (100%) is distributed according to seven eigenvalues. Two empirical criteria are used for selecting the number of axes:

- Elbow criterion: On the eigenvalues scree, we observe a setback (elbow) followed by a regular decay. The axes are selected before the offset.
- Kaiser criterion: Only the axes whose inertia is greater than the average inertia I / v are retained. Kaiser in normalized PCA: $I / v = 1$. We only retain the axes associated with eigenvalues greater than 1.

There is a sharp decline in eigenvalues between the third and fourth ones; the first three components account for 88.2% of the available information. There is a strong “size effect” in our data. Indeed, on one side, the fall of inertia is very important from the fourth axis, which retains only 7% of the total inertia; on the other, the first three axes retain 88.2% of the inertia, which is very good. So:

- Axis$_1$, determined by the T_m factor, carries 53.7% of the available information.
- Axis$_2$, determined by the FW_m factor, carries 20.3% of the available information.
- Axis$_3$, determined by the factor FR_m, carries 14.2% of the available information.

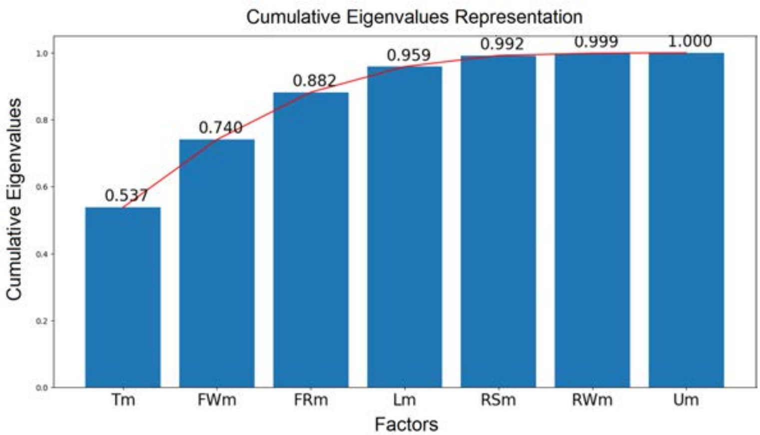

Figure 12. Eigenvalues cumulative inertia.

4.3.5 Results Analysis

Reading for each of the factors selected from the correlations with the seven variables then makes it possible to determine their concrete significance.

In Figure 13 we see correlations on the ellipse; the first component shows the tweet and interactivity indicator. The orthogonal projection of the vectors over the axis shows a positive correlation between the number of tweets, likes, retweets and replies. On the other hand, Axis$_2$ indicates users who tweet and the accounts they follow; hence we notice a strong relationship between users who tweet and the accounts they follow.

The correlations ellipse in Figure 14 also demonstrates that the third component characterizes only followers. The alignment of the variables with the axes shows an independence between certain keywords, so we could have four categories: the first characterized by the number of tweets and the mode of interactivity, a second characterized by the users who tweeted and their subscriptions, a third category characterized by the number of subscribers, and finally a fourth that mixes these characteristics. We can calculate the representation of the quality of the keywords over the axes; to do this, we must first calculate the distances' squares at the origin of the keywords, which also correspond to their contribution in total inertia:

$$CTR_{ik} = \frac{F_{ik}^2}{n * \lambda_k}. \tag{4}$$

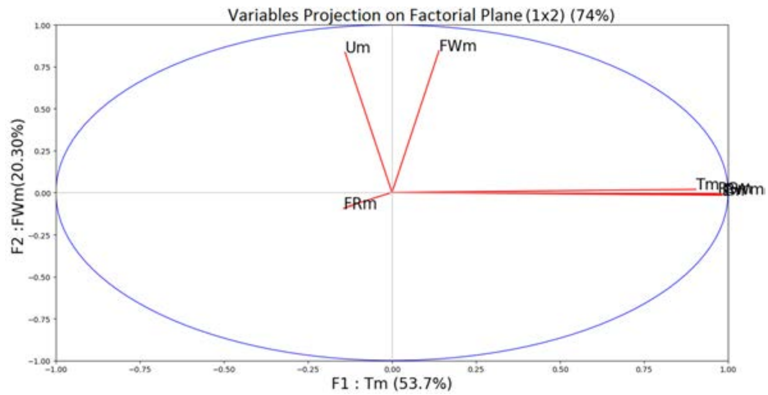

Figure 13. Correlation ellipse ($F1$, $F2$).

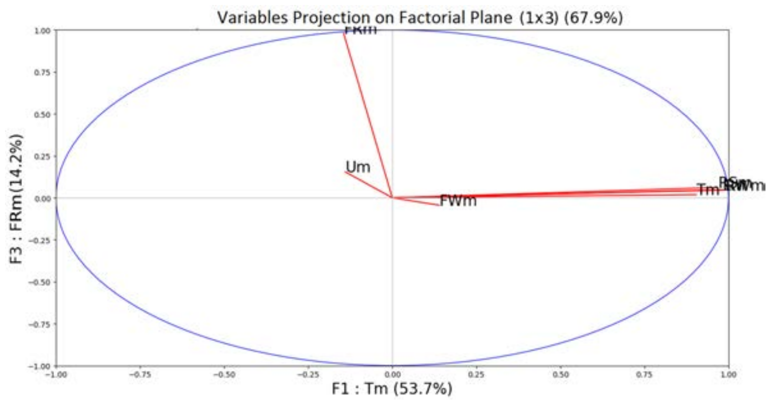

Figure 14. Correlation ellipse ($F1$, $F3$).

To interpret the axis, we need eigenvectors for the variables analysis:

$$CTR_{jk} = \frac{r_{jk}^2}{\lambda_k}. \tag{5}$$

Then, for the interpretation of the variables we use Table 8; we keep the variables whose contribution $|\mathrm{Cor}_x|$ is greater than the average contribution:

$$\frac{1}{\sqrt{p}} = \frac{1}{\sqrt{7}} = 0.378. \tag{6}$$

Regarding individuals, we utilize the keywords correlation matrix and we are interested in individuals with factorial coordinates $|\mathrm{Cor}_x|$ greater than the square root of the first eigenvalue:

$$\sqrt{p} = \sqrt{3.757} = 1.938. \tag{7}$$

The sign gives the meaning of the contribution.

	Axis_1		Axis_2		Axis_3	
	Cor_1	Ctr_1	Cor_2	Ctr_2	Cor_3	Ctr_3
T_m	0.904	0.218	0.018	0	0.019	0
FW_m	0.140	0.005	0.845	0.502	–0.044	0.002
FR_m	–0.145	0.006	–0.092	0.006	0.979	0.965
L_m	0.982	0.257	–0.016	0	0.048	0.002
RS_m	0.968	0.250	–0.005	0	0.062	0.004
RW_m	0.989	0.261	–0.011	0	0.046	0.002
U_m	–0.140	0.005	0.836	0.492	0.154	0.024

Table 8. Eigenvectors.

4.4 Axis_1 Interpretation

- Variables: According to Table 8, we compare the values of the column Cor_1, coordinates of the first factorial axis, to the square root of the average contribution (0.378). We obtain:

–	+
	RW_m
	L_m
	RS_m
	T_m

- Keywords: We use the keywords correlation matrix and we select Cor_1 values for which $|\mathrm{Cor}_1|$ is greater than 1.938; then we obtain:

–	+
	global warming
	climate change
	fossile
	renewable energy

$Axis_1$ does not put any variable in opposition; the four variables RW_m, L_m, RS_m, T_m (number of tweets and different interactivity modes: like, reply, retweet) contribute strongly (98.6%) to the $Axis_1$ formation $(0.261 + 0.257 + 0.218 + 0.250 = 0.986)$. Also it does not put in opposition any of those keywords. According to the column Ctr_1, we can notice that those keywords are well represented $(0.51 + 0.20 + 0.06 + 0.05 = 0.82)$ on this axis: they contribute to 82% of its formation. Those keywords are the most used in tweets (T_m) and have received a large number of retweets, likes, and replies (RW_m, L_m, RS_m). $Axis_1$ demonstrates that keywords global warming, climate change, fossile, renewable energy are the most interesting words to address the environment and renewable energies [19] topic on Twitter and attract the attention of the international community, given the large number of various interaction modes that were received. To validate this result, we can return to the data source (observed values).

Keyword/Variables	RW_m	T_m	L_m	RS_m
global warming	1707789	14660	4397422	311976
climate change	1378278	7867	3520321	157745
fossile	754741	7055	2038534	72665
renewable energy	655451	7297	1607306	64789

4.5 $Axis_2$ Interpretation

- Variables: According to Table 8: we compare the Cor_2 values, the second factorial axis coordinates, to the average contribution square root (0.378). We obtain:

−	+
	FW_m
	U_m

- Keywords: We use the keywords correlation matrix and we take the values of the column Cor_2 greater than 1.938. We obtain:

−	+
	GCAS2018
	electric vehicle
	الطاقة الشمسية
	photovoltaique

$Axis_2$ also does not put any variable in opposition; the two variables U_m and FW_m (number of users and number of accounts following those users) contribute strongly (99.4%) to the formation of this axis $(0.502 + 0.492 = 0.994)$. Therefore the second axis does not put in opposition any of those keywords. According to the column Ctr_2, we can notice that these keywords are well represented $(0.28 + 0.20 + 0.14 + 0.12 = 0.74)$ on this axis: they contribute to 74% of its formation. These keywords are less used and do not gain attention compared to those previously mentioned; they have a wide specific audience: it is the large number of subscribed users who have used them. Axis_2 characterizes keywords used by users with a large number of followers. Using data source values we find, for example, the GCAS2018 keyword is tweeted by 203 users; those users make the keyword be viewed by three million other users.

Keyword/Variables	U_m	FW_m
GCAS2018	203	3035522
electric vehicle	157	1021141
الطاقة الشمسية	97	916254
photovoltaique	40	445964

Table 10. Observed values.

4.6 Axis_3 Interpretation

- Variables: Table 8 is used; we compare the Cor_3 values, the third factorial axis coordinates, with the average contribution square root (0.378). We obtain:

−	+
	FR_m

- Keywords: We use the keywords correlation matrix and we take the values of the column Cor_3 greater than 1.938. We obtain:

−	+
	Global Climate Action Summit

Axis_3 is characterized by a single variable: the number of followers, which contributes significantly to 96.5% of its formation. The third axis characterizes a single keyword: Global Climate Action Summit. According to the column Ctr_3, we can notice that this keyword is well represented on this axis: it contributes to 76% of its formation. This keyword characterizes the keyword that has been noticed by a large number of subscribers, the subscribers of the users. Axis_3 characterizes keywords that have been tweeted by users with a large number of subscribers. To validate this result, we can return to the source data (observed values). Deep research into our dataset about the identity of those users gives us as a result the Donald TRUMP account.

Keyword/Variables	FR_m
Global Climate Action Summit	41981564

Table 11. Observed values.

4.7 Summary and Results

The results are seen by analyzing the matrix of 45 keywords studied on the seven observations variables presented in Figures 15 and 16. The four keywords "global warming," "climate change," "fossile" and "renewable energy" (13, 4, 10, 25) are the words that stand out from the others, and are found at the end of the first factorial axis, which carries 53.7% of the available information. Here, schematically, three groups appear:

- The first (1) groups the keywords characterized by the variables that form the axis. In both Figures 15 and 16, group 1 collects the keywords that are the most used and most influenced; most of these words belong to the category of Anglophones.
- The second (2), according to Figures 15 and 16, groups a mixture of keywords of the different categories. Those of Figure 15 are used by the users having a large number of accounts followed, while those of Figure 16 belong to the users having a large number of followers.
- The third group is in the negative zone of the graph (less than 0). It is a set of keywords that is not particularly characterized by a specific variable [20], but all the variables enter into their characterization.

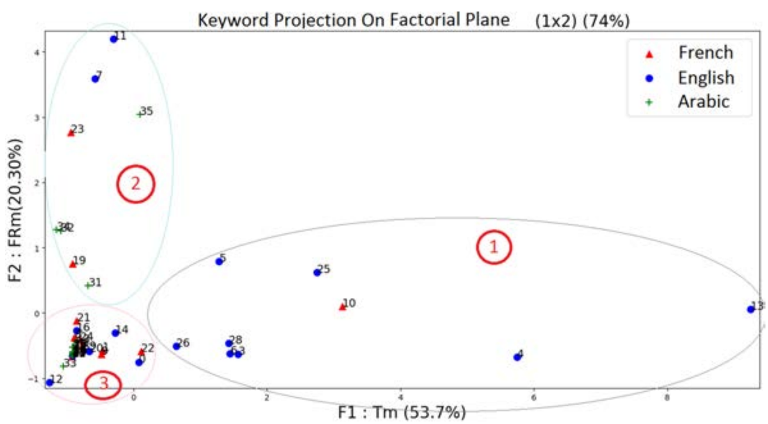

Figure 15. Keywords projection on factorial plane (1×2).

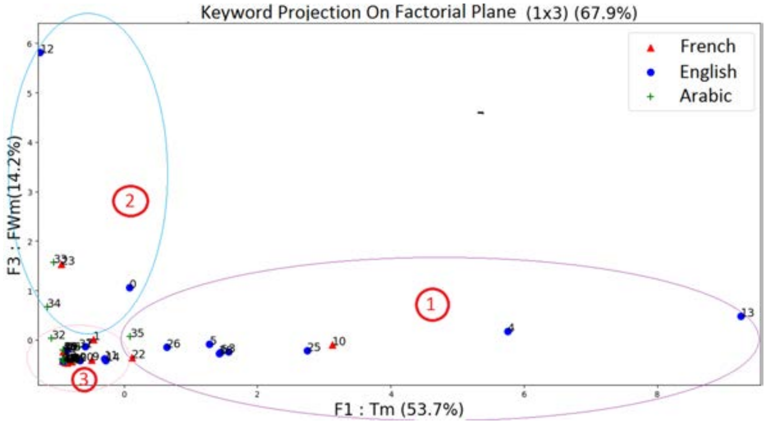

Figure 16. Keywords projection on factorial plane (1×3).

5. Conclusion

In this paper, we assessed the impact of keywords used on the content attraction via the case study of the international community interest on Twitter. We took as an analytic model principal components analysis (PCA), resulting in three factorial axes with defined variables to align with, which allowed us to define categories of keywords. The first is for the most used keywords and different interaction modes. The second is less used and attracts less attention in terms of interactivity compared to the first, and has as audience the users' accounts followed: those keywords produced by a low number of users make them available for a large audience. A third has as audience the users' followers: like the second, a tweet produced by a user is viewed by a large number of followers. It has also been found that the English-speaking terms stand out in comparison with the French-speaking and Arabic-speaking; the words of the latter, apart from الطاقة الشمسية, have weak characteristic measures. To further this work, a study of åthe extracted data content to look for more characteristics could enrich our results. Also, inserting a time axis would be important and allow us to study the impact of time, visualize data evolution and see the keywords' appearance and dissipation dynamism.

References

[1] C. Avin and D. Dayan-Rosenman, "Evolutionary Reputation Games on Social Networks," *Complex Systems*, **17**(3), 2007 pp. 259–277. complex-systems.com/pdf/17-3-4.pdf.

[2] Y. El Moudene, J. Idrais and A. Sabour, "Introduction to Sociology of Online Social Networks in Morocco. Data Acquisition Process: Results and Connectivity Analysis," *First International Conference on Real Time Intelligent Systems (RTIS 2017)* Casablanca, Morocco, 2017 (J. Mizera-Pietraszko, P. Pichappan and L. Mohamed, eds.), Cham: Springer, 2019 pp. 409–418. doi:10.1007/978-3-319-91337-7_36.

[3] J. Idrais, Y. El Moudene and A. Sabour, "Online Social Networks: Study and Validation of the Regular Behaviors of a Targeted Community on a Complex Network Using a Time Series Approach," *2019 4th World Conference on Complex Systems (WCCS)*, Ouarzazate, Morocco, 2019, Piscataway, NJ: IEEE, 2019 pp. 1–7. doi:10.1109/ICoCS.2019.8930747.

[4] J. Hay and D. Flynn, "The Effect of Network Structure on Individual Behavior," *Complex Systems*, **23**(4), 2015 pp. 295–311. doi:10.25088/ComplexSystems.23.4.295.

[5] S. Raychaudhuri, J. M. Stuart and R. B. Altman, "Principal Components Analysis to Summarize Microarray Experiments: Application to Sporulation Time Series," *Pacific Symposium on Biocomputing*, 2000 pp. 455–466. doi:10.1142/9789814447331_0043.

[6] T. Fung and E. LeDrew, "Application of Principal Components Analysis to Change Detection," *Photogrammetric Engineering and Remote Sensing*, **53**(12), 1987 pp. 1649–1658.

[7] C. A. Mertler and R. V. Reinhart, *Advanced and Multivariate Statistical Methods: Practical Application and Interpretation*, 6th ed., New York: Routledge, Taylor & Francis Group, 2017.

[8] M. Kandias, L. Mitrou, V. Stavrou and D. Gritzalis, "Profiling Online Social Networks Users: An Omniopticon Tool," *International Journal of Social Network Mining*, **2**(4), 2017 pp. 293–313. doi:10.1504/IJSNM.2017.10012921.

[9] K. Zhou, C. Fu and S. Yang, "Big Data Driven Smart Energy Management: From Big Data to Big Insights," *Renewable and Sustainable Energy Reviews*, **56**, 2016 pp. 215–225. doi:10.1016/j.rser.2015.11.050.

[10] F. Erlandsson, R. Nia, M. Boldt, H. Johnson and S. F. Wu, "Crawling Online Social Networks," *2015 Second European Network Intelligence Conference*, *(ENIC 2015)* Karlskrona, Sweden, Washington, DC: IEEE Computer Society, 2015 pp. 9–16. doi:10.1109/ENIC.2015.10.

[11] J. Idrais, Y. El Moudene and A. Sabour, "Introduction to Sociology of Moroccan Online Social Networks: Evolution Analysis of the Moroccan Community Activity on Facebook," *First International Conference on Real Time Intelligent Systems (RTIS 2017)* Casablanca, Morocco, 2017 (J. Mizera-Pietraszko, P. Pichappan and L. Mohamed, eds.), Cham: Springer, 2019 pp. 395–408. doi:10.1007/978-3-319-91337-7_35.

[12] E. K. Stigka, J. A. Paravantis and G. K. Mihalakakou, "Social Acceptance of Renewable Energy Sources: A Review of Contingent Valuation Applications," *Renewable and Sustainable Energy Reviews*, **32**, 2014 pp. 100–106. doi:10.1016/j.rser.2013.12.026.

[13] S. Batel and P. Devine-Wright, "Towards a Better Understanding of People's Responses to Renewable Energy Technologies: Insights from Social Representations Theory," *Public Understanding of Science*, **24**(3), 2015 pp. 311–325. doi:10.1177%2F0963662513514165.

[14] A. Petroni and M. Braglia, "Vendor Selection Using Principal Component Analysis," *Journal of Supply Chain Management*, **36**(1), 2000 pp. 63–69. doi:10.1111/j.1745-493X .2000.tb00078.x.

[15] Y. El Moudene, J. Idrais and A. Sabour, "Principal Component Analysis (PCA) Use for Keywords Research Characterization. Case Study: Environment and Renewable Energy Community Interest," *2019 4th World Conference on Complex Systems (WCCS)*, Ouarzazate, Morocco, 2019, Piscataway, NJ: IEEE, 2019 pp. 1–7. doi:10.1109/ICoCS.2019.8930789.

[16] Y. El Moudene, J. Idrais and A. Sabour, "Introduction to Sociology of Online Social Networks in Morocco. Data Acquisition Process: Results and Connectivity Analysis," *International Journal of New Computer Architectures and Their Applications*, 7(2), 2017 pp. 56–61. doi:10.17781/P002330.

[17] I. T. Jolliffe and J. Cadima, "Principal Component Analysis: A Review and Recent Developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **374**(2065), 2016 20150202. doi:10.1098/rsta.2015.0202.

[18] S. Stewart, M. A. Ivy and E. V. Anslyn, "The Use of Principal Component Analysis and Discriminant Analysis in Differential Sensing Routines," *Chemical Society Reviews*, **43**(1), 2014 pp. 70–84. doi:10.1039/C3CS60183H.

[19] G. Dóci, E. Vasileiadou and A. C. Petersen, "Exploring the Transition Potential of Renewable Energy Communities," *Futures*, **66**, 2015 pp. 85–95. doi:10.1016/j.futures.2015.01.002.

[20] C. Bhowmik, S. Bhowmik and A. Ray, "Social Acceptance of Green Energy Determinants Using Principal Component Analysis," *Energy*, **160**, 2018 pp. 1030–1046. doi:10.1016/j.energy.2018.07.093.