Exploiting time series analysis in Twitter to measure a campaign process performance

Eleanna Kafeza
College of Technological Innovation
Zayed University
Abu Dhabi, UAE
Email: eleana.kafeza@zu.ac.ae

Christos Makris, Gerasimos Rompolas
Department of Computer Engineering and Informatics
University of Patras
Patras, Greece
Email: makri, robolas@ceid.upatras.gr

***Abstract*—While there are several metrics to measure business process performance, recently there is an additional requirement from businesses to evaluate business processes based on their impact on users. In this work, we evaluate business process performance using social media analytics. We view a marketing campaign as a business process and we evaluate its performance based on its impact on the Twitter. We propose a new way to calculate the follow relationship in Twitter based on the users reaction to the marketing campaign process activities and we use time series and sentiment analysis for defining and measuring performance. We re-build the Twitter graph based on users reactions to the marketing activities in time and we are using community detection algorithms to identify the size of the follow community and thus we define metrics to calculate the impact of the marketing/campaign process. We evaluate our approach using a dataset for a given politician. We re-construct the campaign process as a set of activities on specific topics (promotions) in time using LDA. Our results show that social media analytics can be used as a valid metric for assessing business processes performance.**

***Index Terms*—marketing campaign business process, time series analysis, community detection, analytics**

I. Introduction

In [1] the authors identify the need to associate the marketing activities within an organization to its business processes.The development and execution of advertising and promotion programs can be considered as sub-processes of the Customer Relationship Management Process. Although until recently marketing activities have not been examined in the context of a business process framework, this has been changing as creating value for the customer is being of paramount importance in the recent competitive environment. For example, as mentioned in [3] in the case of political campaigns, the template based creation of web sites and the commonality on the creation of social media profiles result in a homogeneous image of the politicians. On the other hand, the strategy they follow in the dynamic Twitter activity differentiates them and as a result a lot of attention has been given recently in such marketing activities.

The marketing community has identified and documented the importance of performance measurement of marketing acitivities and its impact on the firm performance, profitability and stock returns [2]. In most cases to assess a campaign's impact on the public, a survey based methodology is followed.

In this paper we are interested to investigate the link of marketing communication processes to its impact to customer satisfaction based on the social media reaction. The establishment of this link is well documented in the literature [8] but our objective is to quantify this connection. To our knowledge this is the first effort towards that direction in systematic and in a fine granularity level. We are focusing our attention to marketing campaigns in Twitter.

This is an interdisciplinary research with several contributions. Initially, we argue that organizations usually have a repeated way of conducting marketing campaigns and this should be represented as a business process. Such representation will also contribute to the integration of the marketing procedure to the rest of the firms' business processes. We give an example of a marketing campaign business process in BPMN. We provide a set of measures for counting the impact of a campaign. We introduce a new type of relationship graph in Twitter: the keepup graph. We propose a metric where the behavior of each node is represented as a multidimensional time series. Each node connection with a neighbor node is defined as the degree that both time series follow the brand account time series. Based on this new type of relationship in the Twitter graph we produce a set of metrics that count the impact of the campaign in time, thus we count the performance as it evolves in time. We define community creation as a main metric for counting the impact of a campaign and we examine how communities change in time.

For example, in a political campaign users that tweet following the politician tweet behavior, can be considered as users that react to the politician's messages. Grouping together such users with similar reactions, we can determine the extend to which the campaign manages to create groups of activated followers.

Our results show that keepup graphs can provide a parametric environment where marketers can get useful insights about the performance of their processes.

The rest of the paper is structured as follows: Section II presents the related work. In section III based on the marketing literature, we represent the steps of every marketing campaign with an organization as a business process. In section IV we model Twitter relationships using time-series analysis, constructing the keep up graph. he performance metrics are

2474-2473/17 $31.00 © 2017 IEEE
DOI 10.1109/SCC.2017.17

explained in sectionV. Our results follow in section VI-D and we conclude with section VII.

II. Related work

In business advertising and in marketing campaigns it is important to identify the target group of people to address. Recently social media have emerged as a platform where users' interests and satisfaction can be extracted based on their online behavior. In existing literature, the problem of measuring the impact of business social media behavior on users has been addressed using computational linguistics and characteristics of social network structure, in a limited manner. In most cases, batch processing of reviews and comments have been used to measure and reaction on users' business experiences [23]. In our approach, we take a different approach and measure the responsiveness of tweeter crowd to the marketing campaign using a variety of metrics. Our metrics are based on the keepup Twitter graph, a graph that is constructed by representing each user behavior as a multidimensional data point in a time series. We are exploiting such time series structures to measure the campaign's performance. For measuring the dissimilarity of time series we are using the Dynamic time warping (DTW) [11] which is a popular and traditional technique for comparing time series, deforming optimally one of the two input time series into the other. This time alignment property is useful for comparing our time series, as users don't interact simultaneously in the social media, nor do they share the same time zone. This technique has been successfully applied in many applications such as speech [13] and gesture recognition [14], but also as a similarity metric for financial multivariate time series [15].

The extraction of communities in a graph has gained the interest of scientists in multiple areas such as graph theory and social network analysis. Community detection is the problem of identifying structures of grouping of nodes which demonstrates high coupling and low cohesion. In [16], Newman proposed an algorithm that uses modularity as the measure of partition quality and pick the partition that would maximize modularity. The modularity-based criterion has been proved to be a meaningful way for identifying the presence of community structures in networks because it quantifies the quality of the divided communities. Dense internal connections within communities and few connections between them is the criterion for deciding communities partition. This criterion is in accordance to the reasoning of the Twitter network where groups of users with strong relationships are interrelated through the "follow" connections.

In this work we are interested in identifying an established community detection algorithm and augment it with users' behavior information towards a company in order to identify communities that are more consistent.

Among the existing approaches for community detection, the maximization of the network modularity approach is well known and widely adopted. Since the problem of maximizing the modularity of the network in NP-complete [17], the most popular method for modularity based community detection has been introduced in [12], the Louvain method. The algorithm that has been characterized as simple, efficient and easy to implement [18], has at least 1000 citations as mentioned in [19] and there is recent ongoing work in extending it [19], [20]. Although there are several other approaches Louvain method is an established way of computing network communities and as such satisfies our objective.

III. Representing the Twitter communication as a business process

In this work we focus on a subset of the aspects of the chain of marketing productivity as presented in [4]. We examine the tactical actions of the Twitter marketing campaign as determined by the promotion strategy and the customer impact. We propose to view the marketing communication performance in a finer granularity level. We represent a marketing campaign as a business process. Our approach for the development of a generic Twitter process, is based on the principle of strategic consistency: in the absence of market specific differences in strategy, organizations will use similar content and communication irrespective of the market. Moreover, the dynamic aspect of a Twitter campaign with the respect to the process structure, can be further modeled with the use of dynamic business process models. In this work, we

Based on the requirements presented in the marketing communications literature we present a generic model of a Twitter process in figure 1. A marketing communication process needs to start specifying the objectives, goals and the specific application of the business strategy through the specific campaign (Strategy implementation). Afterwards a decision needs to be made whether a new account will be created or the basic company's Twitter account will be used. Subsequently there are several alternatives that can be taken based on the strategy implementation document.

As presented in [5] Twitter can be used as an interactive communication medium or as an re-active. When the interactive communication is selected, a set of hashtags, re-tweets, and mentions will be created. In the policy level the percentage of each type determines the degree of interactivity. When the re-active communication is adopted a set of replies should be created. This is presented in the business process so that based on the policy one or both ways can be followed. At this stage the percentage of interactive re-active messages over the total communication activity for the specific instance is calculated. These percentages are determined based on the strategy implementation document. At the next stage, in the preparation activity, the realted content is retrieved and subsequently the actual communication messages are created. In a typical campaign, tweets' creation is based on a set of policies i.e. information sharing, emotion invoking or action invoking as presented in [6] where the Starbucks' twitter marketing communication is discussed. The tweets generation is usually followed by other activities that aim to encourage users to respond. At the same time, replies, mentions, re-tweets could be created. This is an iterative process that can occur until some specified goals have been reached. Several activities

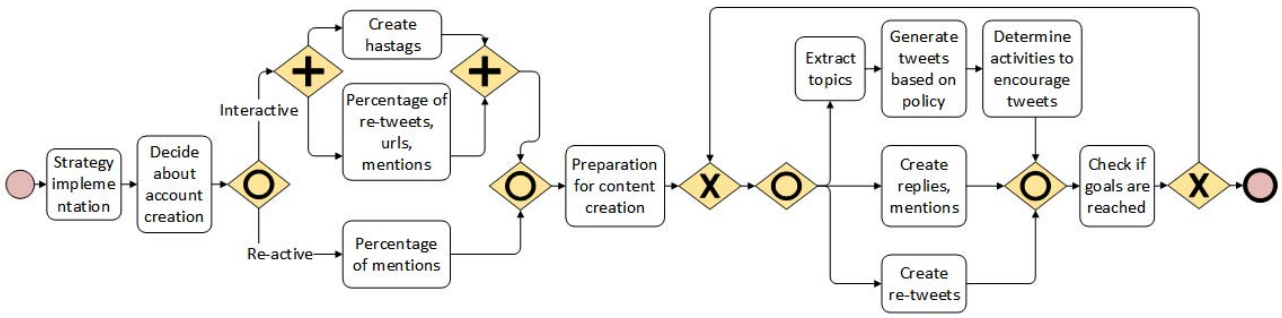


Fig. 1. Twitter Communication Business process

in this process could also be automated. Figure 1 represents one possible representation of a Twitter marketing campaign process.

IV. Modeling the keepup relationship in Twitter

Time, cost, quality and flexibility have been traditionally the main dimensions in measuring performance of business processes and several qualitative and quantitative models have been developed to implement these measures [7]. In the marketing literature it is often sited that a marketing communication process needs to be measured based on the influence on the marketplace. Models are required that can measure how the marketing process impacts the customer behavior [4]. In this context, given a marketing process, we are interested to measure its impact. In this section, we propose a new way of accurately extracting the influence of a marketing communication process, by introducing a Twitter graph where edges represent the impact of the messages.

Our approach can be outlined as follows:

- we represent the social media behavior of a user as a timeseries; this timeseries is formed by statistics on his tweet behavior as it evolves in time,
- we define a distance metric called *keepup* between two users based on the relationship of their time series in comparison to a reference time series, that of the brand name
- we form a user graph where the nodes are the users and the weights of the edges is given by the respective keepup metric,
- we apply on the graph a community detection algorithm and use both the user graph and the communities detected in order to calculate the impact of the marketing/campaign process.

A. The keepup Twitter graph

Given a brand name, the overall current stage of the influence of the brand at a given point in time can be determined by the number of followers. If consider the case of political campaigns, in the literature the number of followers over time has been used as a metric to determine popularity. An example can be found in [3] where the number of followers for every leader for Australia's 2010 election in June-August has been used, to indicate the change of the degree of attention through the month. This change is generally attributed to media coverage.

When looking into the *FOLLOW* relationship in the Twitter graph, one can conclude as described in [9] that Twitter acts as both information and social network. In this paper, we focus on the information network aspect of Twitter and we create a subset of the Twitter graph based on the users common interests as extracted by the hashtags. Such subgraphs are dynamic and are created and processed as an instance associated to marketing business process instantiation. Their objective is to provide measurements for the performance of the business process.

At a given point in time we are interested to extract from the Twitter graph all the tweets that contain the hashtag of the campaign and/or the hashtag of the brand. The unique users that tweet on the hashtags are those that are related to the brand, This is the sub-community of the Twitter graph that we are interested to look into. Algorithm 1 presents the steps of the crawler.

Algorithm 1 Nodes extraction crawler for a time interval

Input: One or more hashtags keywords #keyword
Output: A set N of Twitter nodes that tweeted with the hashtag and their associated tweets
Start at $Start\_date$
Let Tw be an initially empty set
Retrieve in Tw the tweets tw_i that include the #keyword in their text and finish at $Finish\_date$.

```
1: for all tw_i ∈ Tw do
2:    n_i=user who did the tweet tw_i
3:    N=N ∪ n_i
4: end for
```

In existing literature Twitter content has been used to classify users, to do topic analysis, to identify influential users, to identify the change of the relationships. In most cases, the content is extracted and a set of data mining techniques are applied. In this work, we differentiate from our previous approaches [10] where we model user behavior based on the activity of the user on a given corpus of tweets and we examine the user behavior as it changes in time.

We define user behavior of the user n_i at a given point in time t_j as a triple n_{it_j} (x_1,x_2,x_3) where x_1 is the number of tweets, x_2 is the number of re-tweets and x_3 is the sentiment expressed in the tweets and re-tweets of user n_i for the tweets and re-tweets that have been posted at time t_j. In order to count the emotion, we are using an existing metric in R, where the number positive and negative words are measured based on a given dictionary. Any other metric could also be applied.

Given a user n_i let $N_i(n_{it_1}$,...,$n_{it_k})$ be the behavioral data retrieved at consecutive and equal time partitions t_0 to t_k. This time series of triplets, represent how the user behavior changes in time. Note, that this behavior is defined with respect to a specific topic, since tweets are gathered as explained in Algorithm 1 based on a hashtag.

B. Representing the keepup Twitter graph

We introduce a new approach on representing relationships in the Twitter graph. We are interested to measure the impact of the marketing campaign process, therefore we measure the reaction of the users to the brand account main activity (tweets and re-tweets), as it is represented through their time series.

Let N be the users that discuss upon the topics #keyword. Every $n_i \in N$ is represented as a time-series N_i. We differentiate one node time series and we call it N_c. This time series represents the brand account node. Semantically, N_c is the behavior of the brand account during the marketing campaign period.

We define as *distance*(N_c, N_i) a measurement of the disimilarity of the brand name time series N_c and the user time series N_i, normalized to [0,1]. We implement the distance function applying the Dynamic Time Warping (DTW) [11] algorithm, whose rational for computing the distance is to identify that deformation of the time axes of the two time series which brings them as close as possible. The output of the function is the DTW distance i.e. the minimum global dissimilarity.

We introduce a new type of relationship that we call it ”keepup”. The keepup relationship computes the degree in which two nodes n_i and n_j keepup with the brand account node and is defined as follows:

$$keepup(n_i, n_j) = 1 - (|dist(N_c, N_i) - dist(N_c, N_j)|) \quad (1)$$

For example, in Figure 2(a) the left time series represents the N_c, where each line is one of the three dimensions. In Figure 2 (b) there is a time series N_i of a user. When computing their similarity based on the DTW algorithm, we find that they are associated with a keepup value of 0.974. Please note that *dist*(N_c, N_i) computes the distance between N_c and N_i as a measure of dissimilarity between the time series of the brand account n_c and the user account n_i normalized to [0,1]. Since $keepup$ measure the similarity relationship between nodes, we subtract from 1. Initially the Twitter *keepup* graph G is a complete graph where the nodes are the users that have tweeted with #keyword and two nodes are connected with a weighted edge of *keepup*(n_i, n_j) value. Based in the application, we define a threshold parameter th such that for all node pairs in G which have and edge with weight less than th the edge is deleted. This process produces the keepup graph G_{th} (Algorithm 2).

Algorithm 2 Twitter keepup Graph construction

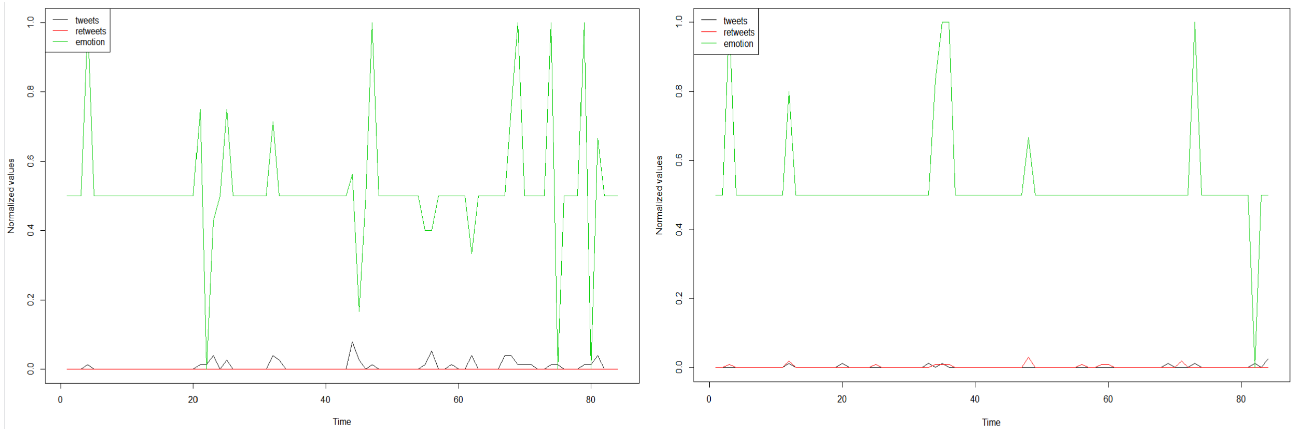
Input: A set of nodes N, a threshold th, a function for the calculation of $distance$
Output: The keepup Graph G_{th} (N, E)

```
E=empty
for all (n_i, n_j) ∈ N do
    keepup(n_i, n_j) = 1 − (|distance(N_c, N_i) − distance(N_c, N_j)|)
    e_ij = keepup(n_i, n_j)
    if e_ij <= th then
        E = E ∪ e_ij
    end if
end for
```

V. Measuring the impact of the campaign

We use two different types of metrics. After the execution of the process instance, following the procedure described in section IV-B, we construct $G_{th}(N,E)$ where N is the number of nodes and E the number of edges. The size of G_{th} in both nodes and edges gives an indication of the campaign impact. Note, that G_{th}, includes the nodes that have tweeted with the #keyword during the campaign and from them, those that kept-up in behavioral attitude with the campaign. As a second metric, we are interested to measure the communities that have been created as a response to the campaign. For this, we use a community detection algorithm to identify the communities. There are several such algorithms, but in our implementation we are using the Louvain algorithm whose objective is to maximize the network modularity [12]. *Communities*(G_{th}) represent the set of communities in G_{th}. We introduce *Quality*(G_{th}) where *Quality* can be a measure the quality of the clustering. In our experiments, given a set of community cluster produced by the Louvain algorithm, we use $Silhouette$ to measure how similar is a user to the users of its own cluster as opposed to other clusters. The similarity is measured based on the average distances of the three-dimentional points. Our approach can be applied in a finer granularity level, and record metrics and their change in time, during the campaign for each community that has been created.That will give insight to the marketers and they can re-design the process of the campaign accordingly if needed.

Let t_0 be the time where the campaign starts. Let t_i be a critical point in the campaign and let G_{tht_ik} be one of the communities of the*keepup* Twitter graph at time t_i. Then, $Metric(G_{tht_ik})$ is a tuple that contains: $Num(n_{t_i})$ the number of its nodes, $Num(e_{t_i})$ the number of its edges, $emotion(G_{tht_i})$ the measurement of the positive emotion in the tweets and re-tweets of time t_i, $Num(tw_{t_i})$ be the number of tweets at point in time t_i, $Num(retw_{t_i})$ the number of re-tweets and $silhouette(t_i)$ the silhouette value for the community. Then for every of the above metrics, we define



(a) Trump

(b) Random user

Fig. 2. Multidimensional Time series

the degree of change between two point in time t_i and t_j as the difference in values. For example the degree_keep_up on the edges dku_e is:

$$dku_e(E_{tht_i}, E_{tht_i}) = Num(e_{t_j}) - Num(e_{t_j})) \quad (2)$$

The degree of keep up, can demonstrate how the impact of the campaign has changed in a per week base. Moreover, we define a general metric

$$Community\_distance(G_{tht_i k}, G_{tht_j k}) = Euclidean\_distance(Metric(G_{tht_i k}, Metric(G_{tht_j k})) \quad (3)$$

VI. Implementation

A. Data description

For our implementation we selected a political marketing campaign and we collected the tweets with the #Trump. Our objective was to measure the users' reaction to the tweets of the Trump tweet account (i.e. brand name account). We collected data from 01/01/2017 until 28/01/2017 and in order to measure the degree of keep up, we created datasets for every week. Although in our future work we intent to apply out methodology for Big Data as well, that is out of the scope of the current paper. Therefore, we implemented a random sampling and we collected 3000 nodes for each week. The following table provides details for our sample.

TABLE I
#Trump dataset

Metrics	Week 1	Week 2	Week 3	Week 4
#Tweets	28221	28961	27424	39474
#Retweets	32160	37448	27434	19131
#Nodes	3000	3000	3000	3000
Nodes % of initial dataset	2.11%	1.16%	0.82%	0.72%

TABLE II
Trump's dataset

Metrics	Week 1	Week 2	Week 3	Week 4
#Tweets	48	40	50	38
#Retweets	0	0	0	0

We retrieved the data applying the Algorithm 1. In every case (i.e. weekly graph, monthly graph) for all the 3000 nodes of our sample we computed the complete graph and calculated all *keepup*(n_i, n_j) values of the edges as depicted in 2. In our experiments we set the parameter threshold th as the average of the maximum and the minimum keepup values:

$$th_{average} = (Max_{keepup} - Min_{keepup})/2 \quad (4)$$

We used R for our implementation. The code of this paper is available at github [1]. Moreover, we used the DTW algorithm as implemented in R, for the distance calculation.

B. Process re-engineering

For out experiments we could not have access to the initial marketing campaign of the brand account therefore we applied reverse engineering to re-construct some possible steps of the campaign. Specially we looked into the problem of re-building the part of the process where tweets are created based on the policy. For our brand account, *realDonaldTrump*, there were only tweets, as presented in the dataset description. Hence it can be inferred that a highly interactive communication approach was adopted. We also examined the Trump web page, in order to assess other activities that might have encouraged tweets, but our results show that the web page topics are not related to the tweets topics therefore the web page was not used as a mean to encourage tweet participation in any of the four weeks.

1. https://github.com/GerasimosR

We used LDA [22], to implement topic extraction for each week in order to draw some conclusions about the policy. We extracted three topics for each week, represented as a set of keywords as extracted and ranked by LDA. For example for the first topic of the first week, issues like the *border, together, intelligence* and *media* were discussed. Moreover, three hashtags have been used for all topics and not embedded URLs. We have assumed that the process iteration is happening every week.

Although we cannot re-construct each activity step, we can decide about some general policies that have been implemented in the process. It is a highly interactive process, there is a variation in topics selection (i.e. mainly there are no common keywords). There is a change in policy for some weeks, where an approach with several hash tags and references (URLs) is adopted .

Having reconstructed the basic policy, we can apply our methodology and measure the performance.

C. *Using communities for measuring Performance*

As mentioned earlier the keepup Graph represent the extend to with the users are keeping up with the brand name process campaign. In the marketing literature, the need for measuring the impact of the marketing campaign has been documented but also there are studies that the performance of the compaign in social media networks depends also on its adjustment to the community norms [21] and increased attention has been paid to the creation of consumer networks. Hence, the communities that have been created as a reaction to a marketing campaign is a valid measure of its performance. Although in our implementation we used the Louvain method for community detection, other approaches can also be adopted based on the needs of the application.

D. *Results*

Table III shows the results for our measurements. In all cases, two users' communities were created and we are tracing the changes comparing the differences in behavior for the bigger and the smaller communities. For all four weeks of the campaign there were two communities created, and the threshold value was determined as presented in equation (3). The lines in the table represent all the different metrics we have developed for counting change and the columns represent the two community graphs (Com1 and Com2) that were created for every week and the size of Com1 is larger than the size of Com2. Hence the first value of the table +7.7% represents that for the $G_{th,week1,Com1}$ and the $G_{th,week2,Com1}$ that represent the biggest communities of the two weeks, in the second week the degree of keep up i.e. $dku_n(N_{th,week1,Com1}, N_{th,week2,Com1})$ has increased by 7.7 %. This translates to the fact that more people were communicating and hence the campaign of the second week had an impact compared to the one of the first week from this respect. Hence if we assume that the process iterates every week, the topics selected in the second week had a positive community impact. We can also conclude that the creation of the second weeks' hashtags (#RepealObamacare, #FakeNews, #lawEnforcementAppreciationDat) that all belong to the first topic as extracted by the LDA (topic 1 keywords: great,intelligence, election, media, now, thank, last, opponents, support, total)were more interesting, attracting more people into the large community. The same holds for the rest of the values of the large community, while the small one, shows a negative reaction. It seems that the keepup users of the brand account are divided into two groups that react differently to the campaign actions. Although the silhouette value of com1 was reduced and of com2 was stable, in absolute number com1 has a silhouette value of 0.67 in week 2 as opposed to 0.5 for com2. Since the silhouette values measure the quality of the communities, we can conclude that the smaller community is composed of users that are not so related in their behavior as in the first community.

A more visualized presentation of the results for each community is presented in Figure 3. It is evident that while on the second week impact of the marketing communication was positive, at the end of week 3 most of the metrics were negative, while things improved in the third week for several metrics. On the other hand, things look different for the samples of the second and smaller communities. There was a positive reaction in week2 until week3, that was changed to neutral or negative in the subsequent week. These figures indicate that there are reasons to reconsider the marketing process and re-evaluate the activities that result to the tweet generation. As a final overall metric we have used the Community Distance that views each community as a multidimensional point whose dimensions are the metrics and calculates the euclidean distance between each two weeks for both communities. Figure 4 shows a graphical representation of the results. We can interpret the meaning of the Community distance as the overall difference in users' behavior from the previous week. From the figure it is evident that overall users' reactions are reducing smoothly for the end of week 3 and more radically for week4. These results can be associated with the goals of the marketing campaign at the specific stage. A reduced users interest is not a positive sign in most cases.

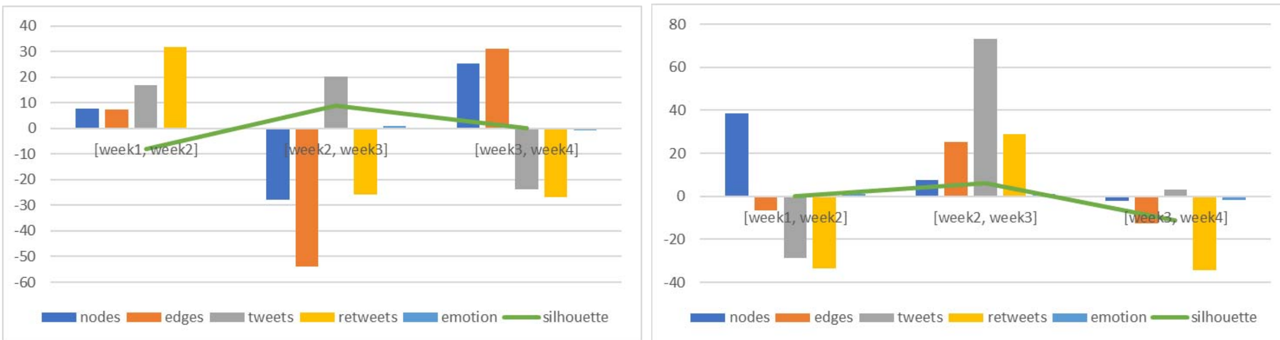
Overall, our results show that our methodology allows for a variety of insights that contribute to evaluate the performance of the instance of the marketing campaign process.

TABLE III
RESULTS

	Week 1 $\rightarrow$ Week 2		**Week 2 $\rightarrow$ Week 3**		**Week 3 $\rightarrow$ Week 4**	
Metrics / Communities	**Com 1**	**Com 2**	**Com 1**	**Com 2**	**Com 1**	**Com 2**
$dku_n(Num(n_i))$	+7.7%	-38.5%	-28.0%	+7.7%	+25.3%	-2.2%
$dku_e(Num(e_i))$	+7.5%	-64.7%	-53.5%	+25.3%	+31.0%	-12.4%
$dku_tw(Num(tw_i))$	+16.8%	-28.8%	+20.4%	+73.1%	-23.7%	+3.2%
$dku_retw(Num(retw_i))$	+31.9%	-33.6%	-25.8%	+28.8%	-26.8%	-34.5%
$Emotion$	-0.2%	+1.2%	+1.0%	+1.2%	-0.9%	-1.6%
$Silhouette$	-8.2%	0.0%	+9.0%	+6.0%	0.0%	-11.3%

VII. CONCLUSION

In this work, we focused at the problem of evaluating business process performance by employing the tool of social



(a) Community 1

(b) Community 2

Fig. 3. Degree of keepup for all metrics (dku)

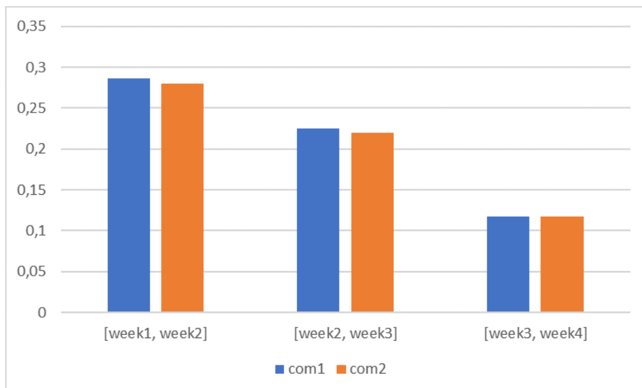


Fig. 4. Overall euclidian distance

media analytics. Our approach entails the definition of a new measure to evaluate the follow relationship in Twitter using the employment of time series, social graph construction based on distance metrics, and community detection algorithms. Experiments performed on datasets concerning a given politician validated that our approach can be fruitfully employed in order to better evaluate business performance; moreover our metrics can be hopefully exploited to other applications engaging social media statistics. As a future work we are considering to transfer and implement our approach in the cloud and to experiment with different and larger datasets; in order to farther validate our results we also plan to fine tune the various components of our approach and provide a parametrization of its various functional components. It could be interesting to embed in our approach more automation and more business formalization in order to more accurately model and reconstruct the marketing business process from Twitter behavior.

REFERENCES

[1] Kristoffersen and S. Sangeeta, *Successful Application of a Customer Relationship Management Program in a Nonprofit Organization*, Journal of Marketing Theory and Practice, vol. 12, no. 2, 2004, pp. 2842.

[2] D. O' Sullivan, and V. A. Andrew *Marketing Performance Measurement Ability and Firm Performance*, Journal of Marketing, vol. 71, no. 2, 2007, pp. 7993.

[3] P. J. Chen,*The New Media and the Campaign*, Julia 2010: The Caretaker Election, edited by Marian Simms and John Wanna, ANU Press, 2012.

[4] R. T.Rust et al. *Measuring Marketing Productivity: Current Knowledge and Future Directions*, Journal of Marketing, vol. 68, no. 4, 2004, pp. 7689.

[5] S. Burton and A. Soboleva *Interactive or reactive? Marketing with Twitter*,(2011), Journal of Consumer Marketing, vol. 28, no. 7, pp491-499.

[6] V. Taecharungroj, (2016), *Starbucks marketing communications strategy on Twitter*, Jounrnal of Marketing Communications, Published online: 10 Feb 2016.

[7] M.Dumas, M. La Rosa, J. Mendling, H. Reijers (2013), *Fundamentals of Business Process Management*, Eds. Springer.

[8] D. W. Vorhies and N. A. Morgan, (2005). *Benchmarking Marketing Capabilities for Sustainable Competitive Advantage*,Journal of Marketing, Vol. 69, No. 1, pp. 80-94.

[9] S. A. Myers, A. Sharma, P. Gupta, and J. Lin, (2014). *Information network or social network?: the structure of the twitter follow graph*, In Proceedings of the 23rd International Conference on World Wide Web (WWW '14 Companion), ACM, New York, NY, USA, 493-498.

[10] E. Kafeza, A. Kanavos, C. Makris and P. Vikatos. (2014). *T-PICE: Twitter Personality based Influential Communities Extraction System*, IEEE International Congress on Big Data.

[11] T. Giorgino. (2009),*Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package*, Journal of Statistical Software, vol. 31, no.7.

[12] V. D. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre. *Fast unfolding of community hierarchies in large networks.*, Journal of Statistical Mechanics: Theory and Experiment,P10008, 2008.

[13] H. Sakoe, S. Chiba. *Dynamic programming algorithm optimization for spoken word recognition.*, IEEE Trans.Acoust. Speech Signal Process, vol. 26, no. 1, 1978.

[14] S. S. Jambhale, A. Khaparde, (2014).*Gesture recognition using DTW & piecewise DTW*, International Conference on Electronics and Communication Systems (ICECS).

[15] S. S. Jambhale, S. Chiba, (2008).*An Efficient Similarity Search For Financial Multivariate Time Series*, 4th International Conference on Wireless Communications, Networking and Mobile Computing.

[16] M. E. Newman. *Fast algorithm for detecting community structure in networks*, Physical Review E, 69(6), 2004.

[17] U. Brandes, D. Delling, M. Gaertler, R. Gorke, M. Hoefer, Z. Nikoloski, and D. Wagner. *On modularity clustering*, IEEE Transactions on Knowledge and Data Engineering, 20(2):172188, 2008.

[18] P. Chopade, J. Zhan, and M. Bikdash. *Node attributes and edge structure for large-scale big data network analytics and community detection*, In International Symposium on Technologies for Homeland Security (HST), pages 1-8, 2015.

[19] H. Lu, M. Halappanavar, A. Kalyanaraman, and S. Choudhury. *Parallel heuristics for scalable community detection. In International Parallel & Distributed Processing Symposium Workshops (IPDPSW)*, pages 13741385, 2014.

[20] X. Que, F. Checconi, F. Petrini, and J. A. Gunnels. *Scalable community detection with the louvain algorithm*, In Parallel and Distributed Processing Symposium (IPDPS), pages 2837, 2015.

[21] R. V. Kozinets, K. Valck, A. C. Wojnicki and S. J. S. Wilner, (2010),*Understanding Word-of-Mouth Marketing in Online Communities*, Journal of Marketing, Vol. 74, No. 2, pp. 71-89.

[22] D. M. Blei, A. Y. Ng, M. I. Jordan, (2003) *Latent Dirichlet Allocation, Journal of Machine Learning Research (JMLR)*, Volume 3, pp. 993-1022.

[23] F.Al-Obeidat, E. Kafeza and Bruce Spencer: *Opinions Sandbox: Turning emotions on topics into actionable analytics*, The 1st EAI International Conference on Emerging Technologies for Developing Countries, AfticaTek 2017, Springer.