



Social Media Conversation Monitoring: Visualize Information Contents of Twitter messages using conversational metrics

Carlo Lipizzi^{1*}, Dante Gama Dessavre¹, Luca Iandoli^{2,1}, and José Emmanuel
Ramirez Marquez^{1,3}

1: School of Systems and Enterprises, Stevens Institute of Technology (USA)

2: Dept. of Industrial Engineering, University of Naples Federico II (Italy)

3: Graduate School, Tecnológico de Monterrey Campus Guadalajara (Mexico)

** corresponding author: clipizzi@stevens.edu*

Abstract

In this paper we present a novel method to extract and visualize actionable information from streams of social media messages, analyzed as conversational elements.

Our method has been applied to over 4 million messages related to more than 35 different events, demonstrating good results identifying conversational patterns.

Keywords: Social media; Conversation Analysis; Visualization; Text mining; Content analysis

1 Introduction

Social media is not only creating new opportunities for companies to interact with customers through online campaigns, but also offers businesses plentiful data that can be mined to better inform marketing initiatives and support a variety of business intelligence applications (Royle & Laing, 2014), (Osatuyi, 2013), (Rud, 2009).

Among the several Social Media channels, the focus of this work is on Twitter due to its increasing popularity (Twitter reached 23% of web users in September '14 ((Duggan, Ellison, Lampe, Lenhart, & Madden, 2015)). Additionally, access to Twitter data is relatively easy with low cost, since most users create their content public. Twitter is increasingly gaining marketers' attention when online users employ this social medium as a backchannel (Dork, Gruen, Williamson, & Carpendale, 2012) to follow ongoing events of general interest such as a TV show (Highfield & Bruns, 2013), emergency events (Hughes & Palen, 2009), or conferences and other live events (Reinhardt, Ebner, Beham, & Costa, 2009).

In the next sections we show that the dominant approaches to Social Media mining are biased towards the analysis of users-generated content rather than of the process through which this content is generated. Our challenge is instead to dig deeper into Twitter streams to observe the emergence of shared meaning in online, large-scale conversations by using a quantitative methodology inspired to ideas and theories from studies in discourse analysis. We extract concept maps from social streams, compute a set of conversational and more traditional online analytics and then create visualizations to display the conversational flows. In this paper we used as case study the Apple Keynote presentation that took place in March 2015, a 90 minutes event focused on the new MacBook and the AppleWatch.

2 Current approaches to social media mining in business applications

Commonly used approaches in social media mining are based on Information Diffusion in Social Networks (Bird, Gourley, Gertz, Devanbu, & Swaminathan, 2006). Another important category of methods is based on sentiment and reputation analysis (Brown, 2012) (Saif, He, & Alani, 2012) or on text mining of the content published on social media site (W. He, Zha, & Li, 2013).

Both social and semantic models presume a pre-existing structure and do not capture effectively how the structure of the judgments and contents emerges and evolves dynamically, thanks to the accumulation and exchanges of opinions expressed by users in backchanneling applications. Dominant approaches do not explicitly treat the nature of online interactions as the conversations they frequently are. This is a shortcoming since conversations exhibit dynamics and features that are not well captured by structural approaches.

In terms of visual representation of the analysis, current solutions are either:

- Data visualization and Statistical analysis addressing cases with high number of participants, with low level of conversation
- Virtual conversation analysis focusing on cases with low number of participants and high level of conversation.

Our analysis – and consequently our visual representations - focuses on cases with large number of participants, with an intermediate level of conversation.

3 Methodology

We expanded the conversational framework detailed in (Lipizzi, Iandoli, & Marquez, 2015) with additional metrics and with a visualization layer. The combined steps are:

- Data collection and Preprocessing
- Concept map Extraction: creating the Semantic and Social structures, performing the Traffic analysis, Integrating the data
- Concept map Enrichment, with additional conversational indicators and with event-specific time labeling
- Data Visualization.

3.1 Data collection and Preprocessing

The methodology and the related tool are applied to a corpus of tweets generated by the public during the Apple Event last March 2015. The tweets collection started at the opening of the event and stopped when it ended. We selected this event because it was very focused, very popular and there is a

wide availability of event documentation. The dataset we used was composed by 46,804 data points, containing tweets in English only.

3.2 Concept map extraction

A concept map was created using a topology based method, as described in (Lipizzi et al., 2015). Using this method, the data was partitioned into m time slices of 2 minutes with each bucket containing an average of 924 tweets, and then 2 structures were created: semantic and relational. We then calculated traffic related metrics and finally integrated all the results.

Semantic Structure. Senders and words were used as nodes. They are related if used in the same tweet. The network resulting is a bipartite one. As an example, the 28th partition (G_{28}) has 1160 nodes and 1108 edges.

Following the above mentioned method (Lipizzi et al., 2015), we extracted from the bipartite networks the words networks. Using the same 28th partition as example, the resulting network G_{28V} is composed by 370 nodes and 109 edges. We then extracted clusters using a combination of k-core and Louvain community detection methods.

Figure 1 represents the community network $D_{28Communities}$, extracted from G_{28V} with $k = 4$, leading to a total of 4 communities. The nodes labels size is based on their degree centrality value

Social Structure. In the social structure, nodes are users U connected one to the other is $u_i \in U$ mentions in his/her tweet other users $u_{x...z}$. Figure 2 is an example of this network.

Traffic Analysis. This is based on the overall number of tweets in each slice and their change during each time interval.

Data Integration. The metrics collected in the above steps were compiled into a single dataset, used to analyze and visualize the whole stream of communication. Statistical steps to improve the quality of the data were also done.



Figure 1: Semantic clusters



Figure 2: Relation network

3.3 Concept map enrichment

To obtain more actionable visualization, we enriched the concept map by time labeling the event that has been analyzed and adding conversational indicators.

Time labeling the event. In events like the one being analyzing, the conversational timeline is defined by the speakers' agenda. Using one of the blogs covering the event (<http://www.techradar.com/us/news/wearables/apple-watch-event-live-blog-march-1287686>), we created a log of the event to time stamp each of the key moments (e.g.: the announcement of the new MacBook).

Adding conversational indicators. To analyze the dynamic evolution of the conversation aspects, we extracted the most relevant word in each cluster/topic and the semantic similarity between each

cluster/topic and the next one. For this task, we used the Natural Language Tool Kit in Python, a lexical database – WordNet – and the Lin similarity.

3.4 Data visualization

Figure 3 below shows the first visualization for the Apple Keynote event over time. The vertical axis refers to the semantic distance between two adjacent bubbles – representing clusters of words/topics –, the size of the bubbles is proportional to the number of tweets in that time slice, and the size of the inner white circle is proportional to the number of cluster/topic in the time slice. The aura on the side of the bubbles represents the sentiment for the specific cluster/topic.

Globes are connected by their semantic similarity. This provides a visual of the evolution of the conversation. In this case, the conversation is fragmented, being the subjects of the conversation determined by the speakers.

From this visualization, it can be seen that the conversation about MacBook is much more focused than the others. This may be because MacBook is an established product, generating less reactions, but more specific conversations.

The Apple watch, that was real core of the event, generated overall less focused conversations, most likely because was a new product whose applications and potential users are relatively unknown.

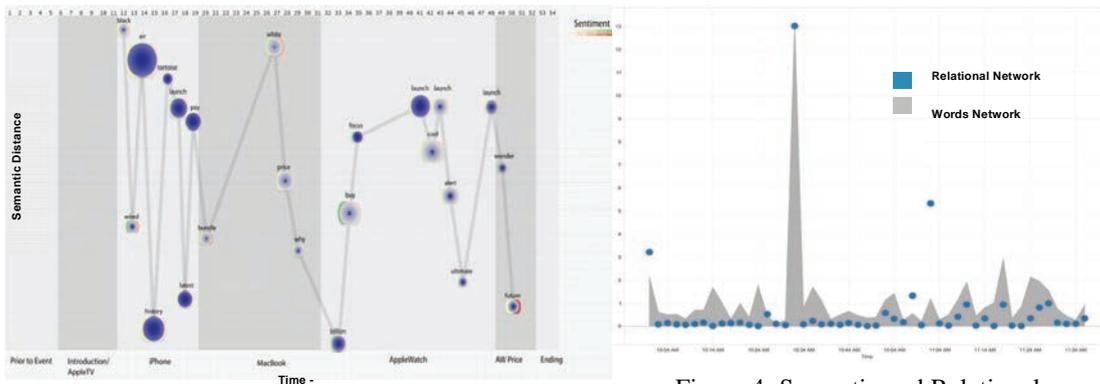


Figure 3: Conversation focus points

Figure 4: Semantic and Relational indicators

To better analyze the reactions to MacBook and AppleWatch, we used the visualization in Figure 4 above, based on clustering coefficient as a measure of homophily. Clustering coefficient gives a measure of how structured networks are. Higher values indicate structured semantic structures in the network of words, while in the relational network indicate more communities, making the subject of the discussion more collective. The highest peak was during the announcement of the new MacBook (around 10:30 AM). This confirms the analysis of the previous visualization. The Mac is a known and appreciated product, so more people can articulate their opinion in a structured conversation, interacting on some of the topics. During the announcement of the AppleWatch (around 11 AM), we had isolated relational peaks with low semantic structures showing relatively highly related people talking about unfocused subjects.

4 Conclusions

Our findings show that our method is able to capture and quantify several features of backchanneling conversational interaction such as: how attention and focus change across the event, identification and measurement of semantic switches in the conversation, exploration VS exploitation oriented conversational dynamics, and evaluation of specific conversational elements.

5 Acknowledgements

The research leading to these results has received funding from the Strategic Research Counsel at the Academy of Finland under grant agreement n:o 293446 – Platform Value Now: Value capturing in the fast emerging platform ecosystems.

References

- Bird, C., Gourley, A., Gertz, M., Devanbu, P., & Swaminathan, A. (2006). Mining Email Social Networks. *The 2006 International Workshop*, 1–7. <http://doi.org/10.1145/1137983.1138016>
- Brown, E. D. (2012). Will twitter make you a better investor? a look at sentiment, user reputation and their effect on the stock market. *Conference Proceedings for the Southern Association for Information Systems SAIS Conference*, 36–42.
- Dork, M., Gruen, D., Williamson, C., & Carpendale, S. (2012). A Visual Backchannel for Large-Scale Events. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), 1129–1138. <http://doi.org/10.1109/TVCG.2010.129>
- Duggan, M., Ellison, N., Lampe, C., Lenhart, A., & Madden, M. (2015). Social Media Update 2014. *PewResearchCenter*.
- He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, 33(3), 464–472.
- Highfield, T., & Bruns, A. (2013). More than a backchannel: Twitter and television. *Journal of Audience Reception Studies*, 1–5.
- Hughes, A. L., & Palen, L. (2009). Twitter adoption and use in mass convergence and emergency events. *International Journal of Emergency Management*, 6(3/4), 248. <http://doi.org/10.1504/IJEM.2009.031564>
- Lipizzi, C., Iandoli, L., & Marquez, J. (2015). Extracting and evaluating conversational patterns in social media: A socio-semantic analysis of customers' reactions to the launch of new products using Twitter *International Journal of Information ...*
- Osatuyi, B. (2013). Information sharing on social media sites. *Computers in Human Behavior*, 29(6), 2622–2631.
- Reinhardt, W., Ebner, M., Beham, G., & Costa, C. (2009). How people are using Twitter during conferences. *Creativity and Innovation ...*
- Royle, J., & Laing, A. (2014). The digital marketing skills gap: Developing a Digital Marketer Model for the communication industries. *International Journal of Information Management*, 34(2), 65–73.
- Rud, O. P. (2009). Business intelligence success factors : tools for aligning your business in the global economy. Hoboken, N.J. : Wiley & Sons.
- Saif, H., He, Y., & Alani, H. (2012). Semantic sentiment analysis of twitter (Vol. Part I , Volume Part I). Presented at the ISWC'12: Proceedings of the 11th international conference on The Semantic Web, Springer-Verlag. http://doi.org/10.1007/978-3-642-35176-1_32