Event-driven TV Programs Web Community Exploration

Ruggero G. Pensa
Department of Computer Science,
University of Turin
Corso Svizzera 185
Turin, Italy 10149
pensa@di.unito.it

Claudio Schifanella
Department of Computer Science,
University of Turin
Corso Svizzera 185
Turin, Italy 10149
schi@di.unito.it

Luca Vignaroli
RAI CRIT - Turin
Via Cavalli 6
Turin, Italy 10149
luca.vignaroli@rai.it

ABSTRACT

With the goal of understanding how major public interest events are perceived by the TV public and how the users' interests evolve in time, we introduce a data collection, integration and analysis framework that allows to compute, characterize and explore dynamic social web communities. We focus on communities of Twitter users that interact each-other around specific TV events and track their public cybersocial activities to analyze and visualize the information diffusion processes and understand how they affect the community structure of the social network.

KEYWORDS

social network analysis, social TV, community detection

Reference format:
Ruggero G. Pensa, Claudio Schifanella, and Luca Vignaroli. 2017. Event-driven TV Programs Web Community Exploration. In *Proceedings of International Workshop on Social Media World Sensors (SIDEWAYS), Prague Czech Republic, July 2017 (Sideways2017),* 4 pages, CEUR-WS.org.

1 INTRODUCTION

People on the Web talk about television. Trending topics and concepts arise from TV users' "cybersocial" activities thus influencing the community structure of the social network. Users aggregate themselves around topics and concepts that emerge from major events. The U.S. presidential primaries have a clear effect on the community structure of users around the world, but such groups are subject to changes as the candidates pronounce some speech in favor or against some policies (e.g., on immigration, protectionism, welfare) that concern them, directly or indirectly. Capturing and understanding the evolution of such communities promptly is of great value for a number of stockholders: advertisers, broadcast programmers, market and society analysts.

Within the RAI research project on the study of the integration between social networks and the TV world, the collaboration with the University of Turin led in the last years to the design and development of a general framework[1] for the collection and analysis of heterogeneous data coming from standard TV sources (EPGs, editorial metadata, audience metrics) and social networks conversations [1]. Within our general purpose social data collection, integration and analysis framework [9], in this paper, with the goal of understanding how major public interest events are perceived by the TV public and how the users' interests evolve in time, we present a new application that allows to compute, characterize and explore dynamics social web communities.

Our application enables both guided and automatic extraction of emergent topics and concepts from Twitter public conversations. Based on such concepts, it builds the social network of users by leveraging their interactions. Two users are considered as interacting each other if they mention each-other explicitly (in a retweet or a reply) or whenever they refer to similar concepts in their conversations. Moreover, our application enables the execution of multiple community detection algorithms [2–4] to uncover the underlying community structure of the network at different time points. Those communities are then described in terms of social influence of their nodes using several metrics (pagerank, betweenness and eigenvector centrality) [7]) as well as in terms of topics and concepts characterizing their interactions (by using summarization). [8]). Finally, the application enables the observation, analysis and characterization of the diffusion of topics and concepts in the overall network and among the different communities [10]. It could then support the activity of researchers, analysts and practitioners interested in social media analysis and network dynamics. Our use case relies on real public data gathered from Twitter, related to one of the most important TV event broadcasted by RAI.

2 RELATED WORK

In the last three years, a number of tools have been proposed to address the problem of complexity and dynamicity in network analysis and visualization (eg., [6, 11, 13]). GalaxyExplorer [6] is an influence-driven visual analysis system for exploring how users influence each other in a social network. The authors use a galaxy-based visual metaphor to simplify the visual complexity of large graphs. In [12], the authors propose a spatial visualization system to detect geo-social event from Twitter conversations. Gazouille [5] is another location-based system for discovering local events in geo-localized social media streams. Insight4News [11], instead, connects news articles to social conversations, by using topic detection and tweet summarization, and performs hashtag recommendation. Finally, in [13] the authors adopts semi-supervised machine learning

[1] http://hdm.di.unito.it/mesoontv.html

Sideways2017, July 2017, Prague Czech Republic

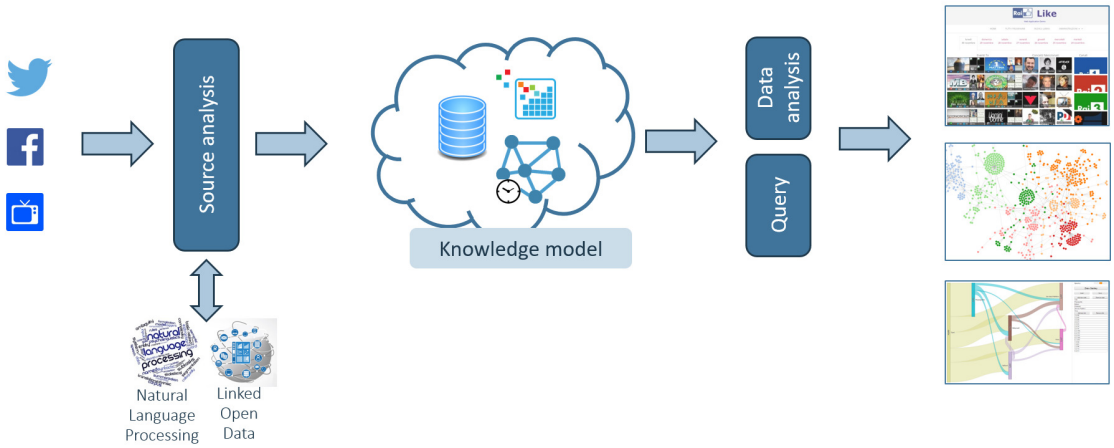

Figure 1: Architecture of the MeSoOnTV framework.

to perform community detection via a label propagation approach that leverages an epidemic spreading model. Differently from them, the focus of our work is *the community*. Our framework, in fact, aims at detecting and characterizing the evolution of the communities gravitating around TV events. Furthermore, we analyze and visualize the information diffusion dynamics at both the single node level and the community level.

3 MESOONTV SYSTEM OVERVIEW

Our framework consists of three internal layers, covering all the phases from data collection, representation and integration to data analysis, and a visualization layer (see Figure 1). More specifically, a **Source processing layer** contains the different modules for collecting all the data from web sources to be conveyed in the social graph and other data representation technologies. It accesses a number of predefined web/social/media sources (Twitter, official web sites, TV channels, ontological information sources) and continuously processes the information collected in real-time to detect the named entities (people, places and events) trough the use of a Named-Entity Recognition module and a topic extraction module. The collected users, topics, concepts and relationships among them are then stored in the **social graph layer** based on Neo4j[2] and MySQL[3], which contains all the modules needed to store and manage the social graph. The **social query and analysis layer** offers functionalities for querying, browsing and analyzing the graph. More specifically the analysis module provides a set of community detection and social network analysis components.

Community detection if performed by adopting two well-known algorithms: *Louvain* [2] and *DEMON* [3]. *Louvain* is a greedy optimization method that attempts to optimize the "modularity" of a partition of the network. Modularity measures the density of links inside communities compared to links between communities [2]. *DEMON*, instead, is an algorithm that detect hierarchical and overlapping communities in networks. It enables the discovery of global communities from multiple *ego-minus-ego* networks thanks to a label propagation algorithm. A *ego-minus-ego* network is defined as the ego-network of a node v without the node v itself [3]. The analysis module implementing *Louvain* and the metrics that measure user influence is based on Gephi[4], while, for *DEMON*, we use our own adaptation of the authors' Java implementation.

The visualization layer is constituted of the web application, that adopt modern web technologies to enable responsiveness and high-level interaction patterns: *node.js* as web page server, a REST API for database querying, *D3.js* as graph visualization and browsing library. To cope with the potential huge amount of data, we adopt a caching mechanism on the layout, nodes, links and communities.

4 DEMONSTRATION SCENARIO

In our use case scenario, we gathered Twitter data from February 9th to 13th 2016. The data are related to the five episodes of the 2016 edition of the *Sanremo Music Festival*, the most popular Italian song contest and awards. Overall, more than $2.52M$ tweets of $176,760$ users have been processed. In these conversations, the relevant concepts are mentioned more than $1.80M$ times, while the overall number of hashtags is $3.60M$. Our goal is to recognize and characterize those communities that gravitates around some specific events that happened during this major national event. In particular, during the 2016 edition, apart from the song contest itself, several events attracted the public interest and were extensively discussed in newspapers, news broadcasting, blogs and social media. For instance, in all episodes, some artists and guests promoted the upcoming law on civil rights for the LGBT community. During the second episode, there was a touching exhibition by Ezio Bosso, an Italian composer known worldwide affected by an autoimmune neurological disorder. In the final episode, Elio e le Storie Tese, a well-known comedy rock band, performed their song dressed as Kiss, the famous metal band; this exhibition was appreciated by Kiss frontman Gene Simmons, thus gaining visibility at international level

A user has the possibility of selecting a set of concepts of interest (e.g., the artists exhibiting during the event, "LGBT", "Ezio Bosso",

[2]http://www.neo4j.com
[3]http://www.mysql.com
[4]https://gephi.org

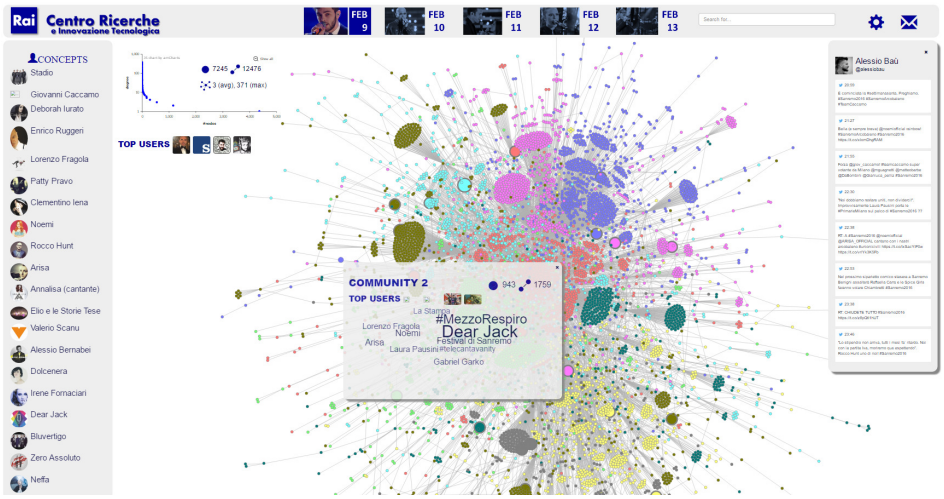

Figure 2: Visualization of community details.

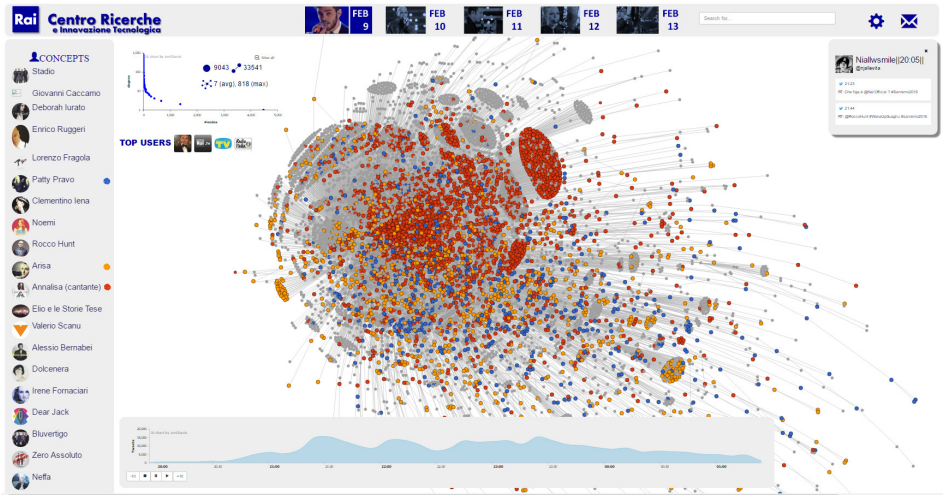

Figure 3: Visualization of concept diffusion.

"Kiss") or a set of "trending/emerging concepts" according to some basic statistics. Then, a community detection algorithm is executed on the Twitter interaction network of the users that mentioned at least one of the selected concepts in their conversations (also including replies and retweets). The network and their communities are visualized as shown in Figure 2. Each community can be then inspected individually and the application returns some related statistics: the list of most influential users (according to pagerank, betweenness centrality or other network metrics) the list of the most relevant concepts and hashtags (identified by a summarization technique) including both selected concepts and other co-occurrent concepts. In Fig. 2, for instance, *Community 2* is characterized by the fact that the involved users talk about Dear Jack (an Italian rock band), while the hashtag *#MezzoRespiro* refers to the title of the song presented during the contest. *Community 5*, instead, is a broad interest community dealing with Ezio Bosso and the civil rights of LGBT people (*#SanremoArcobaleno*).

The application supports also the analysis of the traces of the information diffusion processes involving individual users and communities (see Figure 3). By retracing the timeline of each episode, the application allows the user to visualize and compare the spread of a number of selected concepts in the interaction network. During each time interval, Twitter users that tweet, retweet or reply in conversations mentioning a specific concept c are "turned on" and highlighted with the color identifying c. In this visualization modality, the information diffusion process can be compared with the overall Twitter activity represented by the curve on bottom of Figure 3. This facility supports the discovery of diffusion patterns, as well as the visual comparison of diffusion trends. Notice that basic complex network statistics (number of nodes/edges, number of connected components, clustering coefficient) are always available

on screen. Moreover, the user can observe the time evolution of such metrics referring only to the highlighted subgraph. Though usable on mobile devices, the application is optimized for 4K UHD displays.

Through the described platform, it is possible to identify and visualize phenomena that occur on social networks with reference to the television world such as emerging phenomena, their temporal evolution and the relationships between the entities involved. The conceptual navigation network allows the user to walk through the concepts and involved social users mentioned within television events in a particular timeline based on mentions of mentions during television events.

The flow of data coming from social networks is now seen as a true additional information channel for the end user, for some television programs it is a fruition line for some verse independent of the television program it is generated. The platform described, in addition to being used as a support to marketing choices, can be used as a data source to tell television events from the point of view of users who have lived and commented on them, creating an unseen story that sees as protagonists the viewers themselves with their perception and emotions tried during the viewing of a television program. In fact, Digital Storytelling, or Narrative made with digital instruments, consists of organizing selected content from the web in a coherent system based on a narrative structure, in order to obtain a tale composed of multiple elements of various formats (video, audio, images, texts, maps, etc.). Digital Storytelling is therefore a form of narrative particularly suitable for communicative forms such as journalism, politics, marketing, autobiography, didactics and it could became a new frontiers of the television world to provide information and entertainment experiences that can be visualized through different tools of fruition in synergy with each other.

5 CONCLUSIONS

In this paper we presented an application that allows to find, characterize and explore community of users in social networks conversations. We demonstrate the effectiveness of the approach by exploiting a use case in the TV setting related to a well known Italian song festival. The collaboration between University of Turin and RAI, the Italian broadcaster, is constantly evolving: in the next months, we will integrate other sources that we are already monitoring (e.g., Facebook), as well as new features for a more in-depth characterization and comparison of users and communities.

REFERENCES

[1] Alessio Antonini, Luca Vignaroli, Claudio Schifanella, Ruggero G. Pensa, and Maria Luisa Sapino. 2013. MeSoOnTV: a media and social-driven ontology-based TV knowledge management system. In *Proc. of ACM HT '13*. 208–213.

[2] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *J. Stat. Mech. Theor. Exp.* 2008, 10 (2008), P10008.

[3] Michele Coscia, Giulio Rossetti, Fosca Giannotti, and Dino Pedreschi. 2014. Uncovering Hierarchical and Overlapping Communities with a Local-First Approach. *TKDD* 9, 1 (2014), 6:1–6:27.

[4] Santo Fortunato. 2010. Community detection in graphs. *Physics Reports* 486, 3âĂŞ5 (2010), 75 – 174. https://doi.org/10.1016/j.physrep.2009.11.002

[5] Pierre Houdyer, Albrecht Zimmermann, Mehdi Kaytoue, Marc Plantevit, Joseph Mitchell, and Céline Robardet. 2015. Gazouille: Detecting and Illustrating Local Events from Geolocalized Social Media Streams. In *Proceedings of ECML PKDD 2015, Part III*. 276–280.

[6] Xiaotong Liu, Srinivasan Parthasarathy, Han-Wei Shen, and Yifan Hu. 2015. GalaxyExplorer: Influence-Driven Visual Exploration of Context-Specific Social Media Interactions. In *Proceedings of WWW 2015 - Companion Volume*. 215–218.

[7] Mark E. J. Newman. 2010. *Networks: An Introduction*. Oxford University Press, New York, NY, USA.

[8] Brendan O'Connor, Michel Krieger, and David Ahn. 2010. TweetMotif: Exploratory Search and Topic Summarization for Twitter. In *Proceedings of ICWSM 2010*.

[9] R. G. Pensa, M. L. Sapino, C. Schifanella, and L. Vignaroli. 2016. Leveraging Cross-Domain Social Media Analytics to Understand TV Topics Popularity. *IEEE Computational Intelligence Magazine* 11, 3 (Aug 2016), 10–21. https://doi.org/10.1109/MCI.2016.2572518

[10] Daniel M. Romero, Brendan Meeder, and Jon M. Kleinberg. 2011. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *Proceedings of WWW 2011*. 695–704.

[11] Bichen Shi, Georgiana Ifrim, and Neil Hurley. 2014. Insight4News: Connecting News to Relevant Social Conversations. In *Proceedings of ECML PKDD 2014, Part III*. 473–476.

[12] Shoko Wakamiya, Adam Jatowt, Yukiko Kawai, and Toyokazu Akiyama. 2016. Analyzing Global and Pairwise Collective Spatial Attention for Geo-social Event Detection in Microblogs. In *Proceedings of WWW 2016, Companion Volume*. 263–266.

[13] Ying Wen, Yuanhao Chen, and Xiaolong Deng. 2014. Epdemic spreading model based overlapping community detection. In *Proceedings of ASONAM 2014*. 954–959.