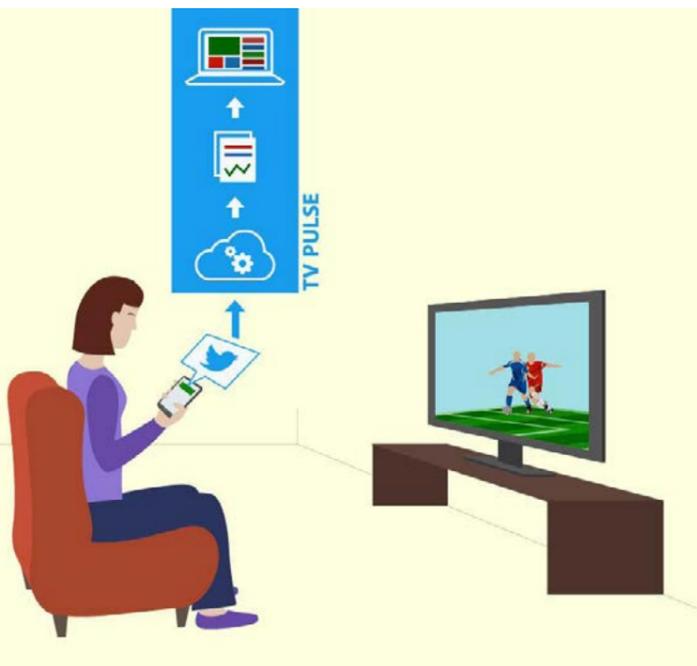# TV Pulse

This project focused on the detection of TV highlights in social networks and on the identification of topics and keywords from those highlights through a semantic analysis of related comments. Those contents were applied on the production of automatic resumes of popular TV programs and on the automatic search of related Web resources (Youtube).

Fig. 1 Matched tweets frequency for the tv program "ídolos" on 12th may 2015. The green diamonds represent the detected events.

Fig. 2 System diagram.

Main Project Team

Diogo Gomes	**TN-Av**
Afonso Vilaça	TN-Av

Funding Agencies

Portugal Telecom Inovação	**20,000€**
Start Date	01/10/2014
Ending Date	01/10/2015

Indicators

Book Chapter	1
Conference papers	2

Two Main Publications

A. Vilaça, M. Antunes, D. Gomes, **TVPulse: detecting TV highlights in Social Networks**, 10th ConfTele 2015 - Conference on Telecommunications, Aveiro, Portugal, September, 2015

A. Vilaça, M. Antunes, D. Gomes, TVPulse: **Improvements on detecting TV highlights in Social Networks using metadata and semantic similarity,** 14ª Conferência sobre Redes de Computadores, Évora, November, 2015
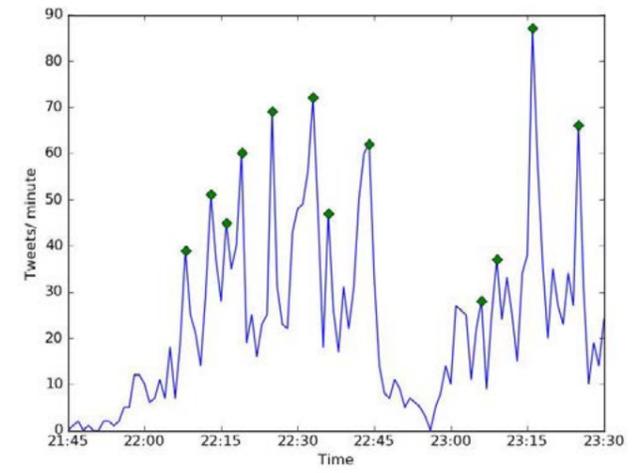
PROJECT WEBPAGE URL
https://atnog.av.it.pt/content/tvpulse
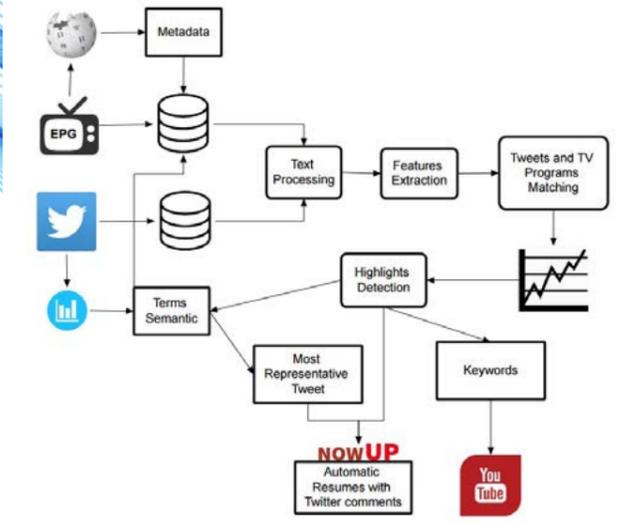
GENERAL MOTIVATION AND OBJECTIVES

Sharing live experiences in social networks is a growing trend. That includes posting comments and sentiments about TV programs. Automatic detection of messages with contents related to TV opens new opportunities for the industry of entertainment information.

Enter here

CHALLENGE

Social networks short messages are a rich source of people opinion and sentiments about what they are living or watching. However, they have a proper syntax and vocabulary, different from natural language. For that reason, information extraction and retrieval from these types of messages are a challenge. Another challenge of this project is on gathering enough rich information to identify TV programs on conversations. The matching between messages and TV programs should be fast enough to be responsive in real type.

WORK DESCRIPTION AND ACHIEVEMENTS

We built a system that is constantly storing information from Twitter and from an electronic programming guide (EPG). The EPG is enriched with metadata gathered from Wikipedia. Tweets are stored in a Cassandra database (for scalability) and TV programs in a MySQL database (for easier information retrieval). Each tweet and TV program is subject to a text processing pipeline from which results a bag of unigrams, bigrams and trigrams. Common identities on both sides are looked for. Another process is the identification of hashtags related to some TV program. For some type of programs regular expressions are also used for detection of some events (e.g. goals in a football match). We focus on two types of events: entertainment shows and football matches. For each type of program, we trained decision trees on annotated data, that were after included in the system. Considering a set of matched attributes, the model relates or not the message with the TV program. In order to detect relevant events, we built an API that identifies tweets minute by minute during a time interval, in a certain TV channel. The detection of relevant events was done using the frequency of related tweets and its second derivative. To understand the topic of an event we used the most used words and the most representative tweets. The most representative tweet is the one with higher node degree in a graph where nodes are tweets and edges are their semantic similarity. An extra feature was added, that meets the goal of finding web resources related to a TV program and its highlights. Using the Youtube Data API, we do automatic video searching using the extracted keywords together with information from the TV program (title and channel). Videos comments are also retrieved, which makes the possibility of enrich the collection of web vocabulary about the program.

We observed that the frequency of the collected tweets varies with the hour of the day and with the day of the week. The time with more activity is the evening, from 6 p.m. to 12 a.m. On weekends there are more tweets, being the Sunday evening the period with more popularity. We also analyzed the most used hashtags. In that top, few were related with TV. Those which were, were mainly about football matches and entertainment shows or contests, typically aired on Sunday evenings. On those programs we successfully detect events. Based on the most representative tweets, it was possible to relate each peak with a moment in the program. The event detection was as successful as the popularity of the program on Twitter. In football matches, goals have a big impact and are always correctly detected. By analyzing several entertainment shows and football matches we proved to have precision values from 80% to 99% and a recall about 50%. We also showed that it is possible to gather more tweets within the same context using semantic similarity. With it we increase the recall and consequently the F-measure. Finally, by observing lists of suggested Youtube videos, we concluded that the automatic search is successful, retrieving related contents. The videos are related to moments of the program (same or other episode) or with a highlighted person.