Project Acronym:	**CATALYST**
Project Full Title:	**Collective Applied Intelligence and Analytics for Social Innovation**
Grant Agreement:	**6611188**
Project Duration:	**24 months (Oct. 2013 - Sept. 2015)**

D3.4 Social network analytics

Deliverable Status:	**Final**
File Name:	**CATALYST_ D3.4.pdf**
Due Date:	**May 2014 (M8)**
Submission Date:	**June 2014 (M9)**
Dissemination Level:	**Public**
Task Leader:	**Wikitalia**

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n°6611188

The CATALYST project consortium is composed of:

SO	Sigma Orionis	France
I4P	Imagination for People	France
OU	The Open University	United Kingdom
UZH	University of Zurich	Switzerland
EN	Euclid Network	United Kingdom
CSCP	Collaborating Centre on Sustainable Consumption and Production	Germany
Purpose	Purpose Europe	United Kingdom
Wikitalia	Wikitalia	Italy

The CATALYST project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n°6611188

Revision Control

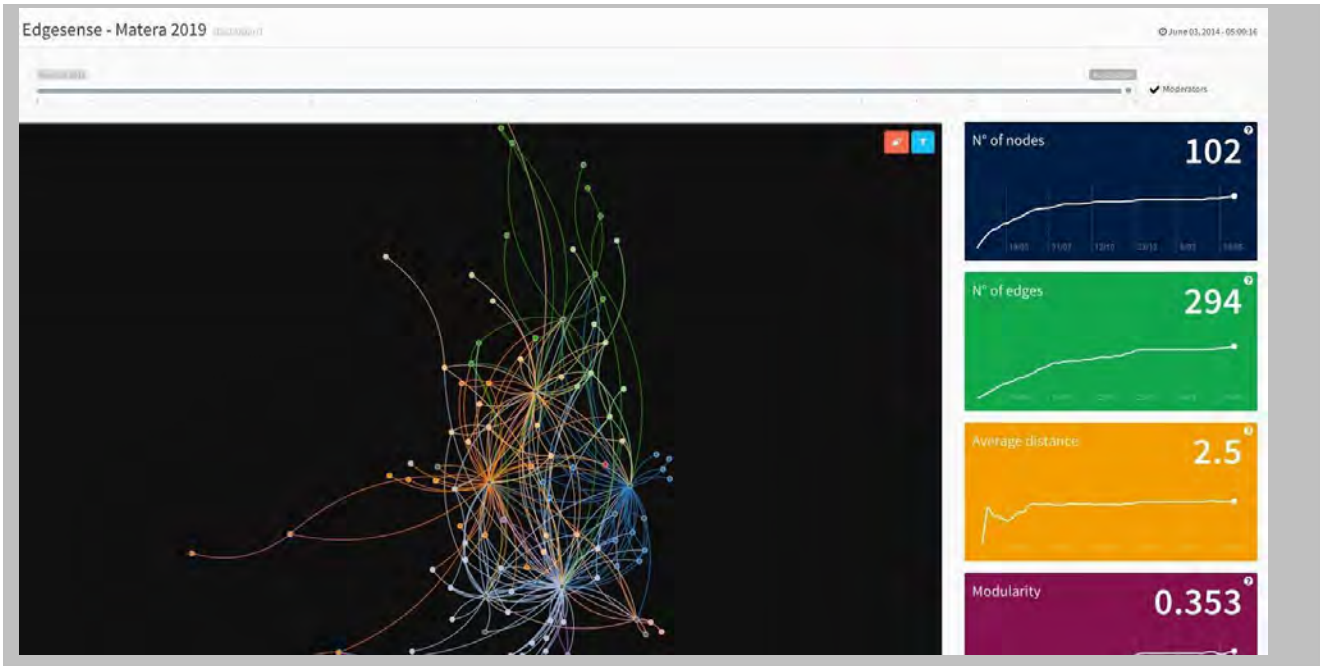
Version	Author	Date	Status
0.1	Alberto Cottica, Wikitalia Luca Mearelli Wikitalia	June, 2014	Draft
1.0	Stéphanie Albiéro	June 12, 2014	Quality check and submission to the EC

The CATALYST project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n°6611188

Table of Contents

The CATALYST project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n°6611188

Introduction



This is a status report on the development of the Edgesense network analysis tool.

The end deliverable is a dashboard which could be used by community moderators to extract insight on the health of their communities. By augmenting online conversations with network analytics, we hope to be able to foster collective intelligence processes.

Edgesense will be implemented first and foremost as a Drupal module that adds social network analytics to Drupal forum and community sites. Nonetheless, most of the code is independent of the underlying Drupal site and is easily adaptable to different platforms.

The CATALYST project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n°6611188

1. Problems solved

- Defining a data acquisition pipeline that is flexible enough and has minimal dependencies with the target platform (this is in fact demonstrated by having being able to link the dashboard to two different community sites without needing to install much on the server)
- Designing a dashboard that can present a collection of metrics and a representation of the network. The network graph can be zoomed, panned and filtered to let the user interactively explore the data. The dashboard controls have been implemented making sure the network view and the metrics are kept in sync with each other when the user navigates, filters and explores the data.

2. Problems to be solved

There are two main roadblocks ahead of us that need to be solved to really have a complete production-ready component:

- Validation of the user experience for the dashboard (does it give the user the information we want to make accessible? Is it able to empower the community manager to gain more insights?). The first testing phase, starting as of June 2014, is expected to yield insights into this.
- Easy installation and customisation of the dashboard for the average community moderator or system administrator is the main challenge.

3. Data model

The script that builds the metrics expects three JSON source files, respectively for: users data, nodes (posts) data, comments data. Their structure is quite simple and the data used is minimal (the script doesn't mind if the files are more complex provided that the basic information is there).

Users data: Each user object should have at least an ID, which is used to identify the user in the other files (UID), the created field that is the date/time (as a timestamp) of the user creation. The administrators are identified as the users that have a role field defined (the script does not look at the content of this file).

Nodes data: (The word “node” is used in the Drupal sense of primary unit of content. Examples of nodes are posts and static pages). Each node is identified by an ID field (NID). The UID field is used to find the node’s author (among the users). The date of creation of the node is taken from the date field. If present, the fields Title and Full text are used to compute the 'weight' of the node and then used in some metrics.

Comments data: Each comment will be identified by an ID field (CID) while the NID field is used to find the node where this comment was placed. When comments are threaded, the ID of the parent comment (PID) is also harvested and used to find identity of the user that authored the parent comment – the target of the interaction. The timestamp field should be the date/time of the comment creation. If present, the Comment and Subject fields are used to compute the 'weight' of the comment and then used in some metrics.

4. Architecture

The current incarnation of the Edgesense SNA tool is made up of two components:

- A Python script used to build the network from the source json files and to compute all the metrics. This script is contained in the python directory (it is build_network.py)

The CATALYST project has received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement n°6611188

- A single-page HTML5/javascript application that reads the json produced by the python script and builds a dashboard with the visualisation of the network and the metrics

The process to populate the dashboard is the following:

1. The views (in Drupal websites, a module called “views” is used to create and store SQL queries) to dump the JSON for the users, nodes and comments are created in the community sites.
2. The build_network script is scheduled to be run at regular intervals, the paths or URLs to the json files are passed to the script.
3. The configuration.json file is created with the analytics tracking id, and the dashboard name to use.
4. When a user opens the dashboard page the javascript code in the page populates the visualization:
   a) The configuration.json file is read from the server and used to set the dashboard title and activate the analytics tracking.
   b) The help.json file is read from the server and used to populate the dashboard contextual help.
   c) The latest.json file is read from the server and, with the information provided there, the processed data is read.
   d) The json containing the processed data is used to set up the dates range to be shown and the latest metrics are shown.
   e) The json containing the processed data is used to populate the network graph.
   f) All the dashboard controls (filters, time slider,...) are activated.

When the dashboard is running, each user interaction with the time slider or with the filters results in the following events:

1. The current view-date is updated.
2. The metrics corresponding to the current view-date are extracted from the processed data.
3. The network is updated by showing only the elements that are valid at the current view-date selected.
4. The graphs are updated to show the current metrics.

5. Code

The code and documentation are to be found here: https://github.com/Wikitalia/edgesense

6. Live instances

- A smaller community: http://matera2019.edgesense.spazidigitali.com/
- A larger one: http://edgeryders.edgesense.spazidigitali.com/

These are both live installations, automatically updated on a daily basis.

The CATALYST project has received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement n°6611188