



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)**

D4.1 Social Network Analytics

Deliverable Status: **Final**
File Name: **CATALYST_D4.2.pdf**
Due Date: **August 2015 (M23)**
Submission Date: **September 2015 (M24)**
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
CSCP	Collaborating Centre on Sustainable Consumption and Production	Germany
Purpose	Purpose Europe	United Kingdom
Wikitalia	Wikitalia	Italy

Disclaimer

All intellectual property rights are owned by the CATALYST consortium members and are protected by the applicable laws. Except where otherwise specified, all document contents are: "© CATALYST Project - All rights reserved". Reproduction is not authorised without prior written agreement.

All CATALYST consortium members have agreed to full publication of this document. The commercial use of any information contained in this document may require a license from the owner of that information.

All CATALYST consortium members are also committed to publish accurate and up to date information and take the greatest care to do so. However, the CATALYST consortium members cannot accept liability for any inaccuracies or omissions nor do they accept liability for any direct, indirect, special, consequential or other losses or damages of any kind arising out of the use of this information.

Revision Control

Version	Author	Date	Status
0.1	Luca Marelli, Wikitalia	September 2, 2015	Initial Draft
0.2	Alberto Cottica, Wikitalia	September 10, 2015	Final Draft
0.3	Lee-Sean Huang, Purpose	September 21, 2015	Quality Check
0.4	Marta Arniani, Sigma Orionis	September 22, 2015	Final Draft reviewed
1.0	Marta Arniani, Sigma Orionis	September 24, 2015	Submission to the EC

Table of Contents

Executive summary	5
Introduction	6
1. Social Network Analysis	7
1.1 Status of the last deliverable and developments anticipated at this date	7
1.2 Test framework	7
1.2.1 How do subcommunities develop and grow?.....	7
1.2.2 Users' life-cycles	7
1.2.3 Community health and other questions.....	8
1.3 User tests	8
1.3.1 A concrete example (Edgeryders / LOTE4).....	8
1.4 Final feedback	9
1.5 Live instances	9
1.6 Evolution of the tool	10
List of Tables and Figures	11

Executive summary

The present document is a deliverable of the CATALYST project, funded by the European Commission's Directorate-General for Communications Networks, Content & Technology (DG CONNECT), under its 7th EU Framework Programme for Research and Technological Development (FP7).

The report details the returns on experience (REX) from the testing cycle of Social Network Analytics in Wikitalia, with formal recommendations on improvements for future testing cycles.

Introduction

For reference, we republish the description for this module set forth at the start of the project:

Wikitalia has developed a processing pipeline as a set of python scripts which using the analytics library networkx. The result is a series of JSON files containing a description of the network and all the computed metrics.

The visualizations have been implemented using web-based technologies. A comprehensive dashboard was implemented as an html5/javascript application to present a subset of the metrics to the users. Integrated in the dashboard is an interactive tutorial which purpose is to learn how to interpret social network analytics. This will help drive adoption and ease the new users experience.

Edgesense also implements a Drupal module that adds social network analytics to Drupal forum and community sites, and is already being adopted by various community sites. Nonetheless most of the code is independent of the underlying Drupal site and is easily adaptable to different platforms, indeed it has been used to analyse communities gathering around mailing lists or discussing on twitter, and it has been extended to integrate in the Catalyst dashboard. Examples of each one of these are included in a later section of the document.

The end deliverable is a dashboard that 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.

1. Social Network Analysis

1.1 Status of the last deliverable and developments anticipated at this date

We have a working integration with the other CATALYST tools via parsing of the CIF file format to enable the social network analytics dashboard to be more useful with the collective intelligence tools developed and make it integratable with the CATALYST dashboard.

We planned to improve the installation process for the drupal module by streamlining the setup of the module making the process to obtain the first dashboard easier. The work on the Drupal module is almost finished and therefore soon we will be able to start the process to get the module approved as an official Drupal module. This will also improve the user experience of installing the module.

For the dashboard, we have planned and started developing improvements to the usability of the network visualization to cover the cases where the current network graph would be too dense to understand completely. This was a consequence of the user testing with real communities and of the feedback gathered after the tests.

The focus of the test is to find out which information about community structure (as captured by network analytics) is most meaningful to community managers and individual users; and how possessing such information affects their behaviour.

1.2 Test framework

Wikitalia has introduced the tool in various online communities, in particular one designed to foster collaboration inside the public administration (InnovatoriPA), and another geared toward for citizen and youth action (Edgeryders). The tests have exposed users to a dashboard that collects metrics and visualizations of the social network defined by the user interactions in the communities (writing content, commenting, etc.).

The tests have been guided by attempting to answer to the following questions about the communities:

1.2.1 *How do subcommunities develop and grow?*

- Who are the central non-moderator members? (Those that are central represent users that could be appointed as moderators in the community)
- How central members add specialization to conversations? What is the nature of the non-moderator interactions between sub communities?
- What, if any, is the influence of moderators in the evolution of sub-communities? If you remove moderators them from the graph, do sub-communities still exist?
- What is the sustainability of conversations in sub communities over time?
- What is the ratio of new conversations being built on the old connections that already exists?
- Excluding the moderators, what is the average degree (in + out) of the different nodes?

1.2.2 *Users' life-cycles*

- Can we predict from early behaviour which users have a chance to become vary active community members aka "tribe elders"?
- What converts a lurker into an active user?
- Is there a typical time window from user creation during which either the new user becomes active (for some value of "active" or is lost forever?

1.2.3 **Community health and other questions**

- Can we maybe break down or build a model for a minimally healthy community?
- Are there reliable tests for a debate's good health?
- Why is debate fruitful and creative in some contexts, sterile and conflictual in others?
- Can we develop a metric as to the optimal proportion of content in a healthy community? Are healthier communities those with a greater share of contributors content than moderators content, or vice versa?
- Do people in more active communities talk to more people? Or do they talk more to the same average number of people?
- Can edgesense be developed to predict different scenarios of how the community could grow?

The tests have been followed by qualitative feedback through a user focus group and survey, which have helped us gather data on the utility of the tool and suggestions for improvements of its usability.

1.3 **User tests**

Over a period of time the users have been shown the dashboard that included a variety of indicators, which have been visualized over a timeline so that not only the community's network characteristics are visible, but also it was possible to identify its changes over time as well.

Usage within Wikitalia and associated communities clearly shows that different community dynamics are captured by the visualization, and initial usage of the tool supports the hypothesis that these visualizations can help interpret those dynamics.

Looking at the network with or without the moderators one can spot some nodes taking some centrality positions, some of them are still active while others left the conversations long ago.

The colouring of the partitions obtained with the Louvain algorithm can help identify subcommunities that have been developing inside the larger online community. Colouring also makes also easier to spot sub-sets of users that are not communicating with each other.

For older communities moving with the slider from the start till now gives a very nice idea on how the community developed passing through different phases.

It has been noticed that whereas the first batch of questions above can be answered to a great extent using only the Edgesense install on a single community, the ones about community health would greatly benefit from a comparative analysis among a varied set of communities. This is especially true when trying to answer to the question whether a community is healthy or not (e.g. it's been noticed that active communities are not the same as healthy ones.)

1.3.1 **A concrete example (Edgeriders / LOTE4)**

Edgeriders have developed over the course of years from an initial community to the current one. It is clear via Edgesense how the old community "gave birth" to the smaller community active now in LOTE4 and UnMon. It shows also how a community (STF) wasn't completely integrated with the community gathered around the LOTE4 event for example (again, it is easy to spot the members who joined the discussions).

Zooming in on the platform from May 18 to November 18 2014 (during the LOTE4 event.)

- Number of edges increased from 3278 to 4003 creating 725 new edges (22 % increase).
- Number of comments increased from 10498 to 13921 (3423 new comments = 32 % increase).
- Modularity changed from 0.313 to 0.346

The network expanded during this period of time. This included different interactions between moderators and members. Also interesting is the moderator's activity from this period:

Table 1. Moderators' activity

Account	May 18	November 18
Natalia	----	in 35 out 17 bet 0.0048
Nadia	in 138 out 160 bet 0.1063	in 193 out 194 bet 0.1337
Noemi	in 144 out 230 bet 0.1545	in 191 out 292 bet 0.1903
Alberto	in 204 out 310 bet 0.2822	in 233 out 368 bet 0.2834
Matthias	in 92 out 101 bet 0.0416	in 114 out 130 bet 0.0491

Not all interactions are LOTE4 related, since most community moderators are involved in different projects plus some community management tasks. The exception is Natalia, as the node was inactive until she took the responsibility to coordinate LOTE4, which included a lot of different tasks. This is clearly reflected by the changes in her metrics as reported by Edgesense.

1.4 Final feedback

These are the main results from the user testing with real communities of different sizes:

- The visual complexity of the network is a challenge in the beginning but Edgesense provides an easy way for users to have more of a network thinking about their communities.
- It provides an overall understanding of the network, how big it is, how modular, and how information flows throughout the network.
- For older and big communities, it is hard to know which are the old members and which are new, and how many of the old members are still active in discussions.
- It provides a quick tool in knowing more the central members of the community and the load of communication done by moderators.
- It shows how the community develops, which could be used in various ways.
- It shows how sustainable the sub-communities are with or without the moderators. (Sustainability here is about keeping the discussions going.)
- Overall, the tutorial helps a lot new users.

1.5 Live instances

There have been a number of installations done during the testing phase that have given us a good feedback on the system implemented. The technology developed is flexible enough so as to easily adapt the processing code and the visualization to extract and get insights from a very different context. Here are some of them, organized by type of community (from a technical standpoint):

Drupal-hosted communities:

- <http://edgesense.innovatoripa.it/>
- <http://wikitalia.spazidigitali.com/edgeryders>
- <http://wikitalia.spazidigitali.com/matera2019>

Twitter conversations:

- <http://wikitalia.spazidigitali.com/mt2019/>

Mailing lists:

- http://wikitalia.spazidigitali.com/animation_fr
- <http://wikitalia.spazidigitali.com/cyou1>
- <http://wikitalia.spazidigitali.com/correspondants>
- <http://wikitalia.spazidigitali.com/espace-sante>
- <http://wikitalia.spazidigitali.com/fablab-fr>
- <http://wikitalia.spazidigitali.com/python>

Catalyst CIF embed:

- http://discussions.blunove.com/edgesense/catalyst_embed.html?url=https%3A%2F%2Fitemap.net%2Fapi%2Fconversations%2F1371081761300912288001398442436%2F%3Fid%3D137108145250626041001433233067

1.6 Evolution of the tool

Before the end of Catalyst we foresee being able to significantly improve the usability of both the network visualization in the dashboard and the install process for the Drupal module.

This is a list of the items that we are working on, which have been directly influenced by the user testing:

- Improve Drupal module installation process and onboarding user experience
- Improve the display for larger networks
- Find a way to distinguish visually between more active users and less active ones
- Enable analysis to concentrate on last months of a community
- Improve user search and highlight
- Enhance colour usage in the dashboard
- Implement an alert for old data
- Implement centrality-based and degree-based colour coding
- Improve drupal module installation process and onboarding user experience
- Build an instructable that explains Edgesense installation/configuration (with examples for Drupal, Tweets, Mailing list)

List of Tables and Figures

Table 1. Moderators' activity.....	9
------------------------------------	---