CATALYST REPORT
DebateHub and Collective Intelligence Dashboard Testing
University of Naples Federico II

1. Overview: Information about Community Partner Involved

1.1. Resumé

This test was carried out by the University of Naples Federico II, Italy– Department of Industrial Engineering, which was selected as a Community Partner in order to test the DebateHub and the Collective Intelligence Dashboard CATALYST tools.

The contact persons involved in this experiment were Professor Luca Iandoli and Dr. Ivana Quinto.

In particular, **Luca Iandoli** is Professor of Engineering Management at the University of Naples Federico II (Italy) – Department of Industrial Engineering - and at the Stevens Institute of Technology (USA), and a former Fulbright Fellow at the Center for Collective Intelligence, MIT. His research activities focus in the following areas: Knowledge Management, Organizational Learning, Collaborative technologies in organizations and networks of small firms. He has published many papers on the analysis of collaborative dynamics in firms' networks through computational methodologies (agent-based modeling, fuzzy logic, social network analysis). He was President for 2011/2012 of the European Council for Small Business and entrepreneurship (ECSB), the largest European academic association on entrepreneurship and small business management research. Currently, he is President for Development of the International Council for Small Business and Entrepreneurship (ICSB).

Ivana Quinto received her PhD in Science and Technology Management at the University of Naples Federico II in 2012. Currently, she is Research Fellow at the Department of Industrial Engineering, University of Naples Federico II. Her research activities focus mainly on how to support decision makers to harness collective intelligence and online collaboration for distributed problem solving and online deliberation processes. Currently, her research interests include Open Innovation, collaborative dynamics in SMEs' networks and organizational learning.

1.2 Motivation

The main interest of this partner to participate to the CATALYST Open Call for Collaboration was ***to test innovative web-based collaborative technologies able to harness Collective Intelligence and online mass collaboration for supporting collaborative and distributed deliberation processes.***

In particular, this research group was interested in testing the hypothesis that: knowledge articulation and visualization through argument mapping formalisms supports critical

thinking and exploration of the problem space if integrated with additional widgets (e.g. visualization tools) to promote social engagement, mutual understanding and better collaborative performance.

In a nutshell, the aim of the collaborative research carried out in this test was to improve collaboration performances and outcomes of online deliberation processes by mediating users' interaction with innovative online collaborative knowledge mapping technologies.

To this end, the test of the Debate Hub and the Collective Intelligence Dashboard was an unique opportunity to better verify the above mentioned hypotheses and to provide some key insights and guidelines for the design of better collaborative platforms and web-based technologies.

2. Experimentation's Goals and Outcomes

The aim of this experiment was to evaluate the performance of Debate Hub, and some of the visualizations that are part of the Collective Intelligence Dashboard, in terms of: knowledge accumulation and organization, as well as users' participation and satisfaction. In particular, we aim at providing users' with some extra information through two classes of visualizations:

- *Social awareness visualization:* is aimed at making users more knowledgeable about the social landscape of the virtual community. In particular, through this class of visualization we provide users with some extra information that could help them to have visibility over the social dynamics and landscape of the community. This visibility is supposed to increase participants' social awareness and accountability as well as to provide social translucence of participants' online behaviour.
- *Content visualization*: should facilitate users in making sense of online discussion and identifying points in the debate where their contributions could be more effectively placed.

More in detail, this field test allowed us to:

1) evaluate how different kinds of visualization impact on users' performance. We measured users performance by calculating users' mutual understanding, perceived quality of collaboration (PEC), perceived ease of use (PEU) and accuracy of individual prediction and tested how the use of different visualisations affect each of these variables.
2) to assess if there is any interaction and synergy between social and content visualization when provided together.

Finally, the test results have been used as feedback to identify improvements and revisions in the design of the Catalyst tools (DebateHub and The CI Dashboard Visualisations we tested).

3. Subjects and Domain Task

An online community of 143 subjects was, initially, involved in a two-factor, synchronous and distributed deliberation experiment. Eventually, only 140 students completed all the activities. The subjects were recruited among students in the same class from an undergraduate program (Economics course) in Industrial Engineering, age 19-22, 61% male. Participation was voluntary and all participants were compensated with some extra academic credits. On the basis of students' activity, three best participants have been selected and will be awarded with some gadgets at the end of University course (around 15th of June).

The subjects were asked to discuss what would be the future price of the crude oil in the short term (3 months from the end of the discussion) and make an individual forecast of its price after the discussion was closed. The criteria used to select the topic are:

- *Realism and information richness*: participants are meant to face a real-world decision task rather than an abstract choice problem, as those used in many lab experiments. Moreover, the task needs to be controversial enough to produce an adequate level of discussion and a variety of positions around the topic.
- *Acceptable task difficulty*: task difficulty needs to be aligned with the skills and background of the participants.
- *Measurability of the outcome:* a task with a unique correct solution allows us to test for the accuracy of individual predictions after being exposed to the debate. In other words, the problem is real and the "right" solution exists, although it is not obvious.

Therefore, following the above-mentioned criteria, we selected as discussion topic the forecasting of the trend price of the Crude oil and it was proposed to the students through the following framing question: ''What do you think will be the trend of the crude oil in the short term (three months)?".

4. Field test design

As the aim of this field test is to compare the performance of different versions of the Debate Hub, as well as the impact of different visualization (i.e. social and content awareness) on users' performance, we adopted a 2X2 experimental design in which one factor is the availability of social awareness visualization and the other the availability of the content visualization. Therefore, participants were randomly divided in four groups, each composed of about 35 students and they were asked to discuss about the price trend of the Crude Oil in the future three months. These students were randomly assigned to the four groups (T1, T2, T3, T4) in order to ensure the validity and universality of our outcomes. In order to verify that there were no relevant pre-existing differences among groups, one-way Anova was performed. We used an *academic proficiency indicator* to test the uniformity of the groups and, thus, to avoid having groups with different characteristics that could produce undesired biases on the results. The formula to calculate this academic proficiency indicator is:

$$\frac{(Average\_marks \ * \ Number\_of \ \_passed \ \_exams)}{(year \ of \ study)}$$

We decided to use this academic proficiency indicator because we supposed that this students' performance can affect final output, given the nature of the task. Instead, there are no reasons to suspect undesired effects attributable to other demographic variables (i.e. age, gender, place of birth), as they should not have a direct impact on users' final performances. In other words, this indicator is a proper discriminant variable respect to it we can test eventually differences among students. After the randomization process, the four groups were made up as showed in the Table 1. We checked that there were not significant differences among the groups by computing an Anova test.

Group	N° of participants	Mean	Std. Deviation
A	36	73.48	21.64
B	36	73.14	25.19
C	35	73.25	30.13
D	36	73.54	27.56

Table 1: Students' University Performance Indicator

Consequently, four different types of instantiations of Debate Hub were developed. Specifically, group T1 used a plain version, i.e. a version of Debate Hub in which all the available data visualization features were removed; group T2 used a platform able to provide only social awareness visualization, that is Social Network and Users' Activity Analytics; group T3 used a platform able to provide only content widgets, that is Debate Network and Activity's analytics; finally, group T4 used a platform able to provide both social and content visualization, namely Social Network and Activity's analytics. Figure 1 shows what visualization each platform provides while Figure 2 shows the type of Visualisations tested.

		CONTENT	
		NOT	*YES*
SOCIAL	*NOT*	**T1**	**T3**
	YES	**T2**	**T4**

Figure 1: Groups/platforms and types of provided visualization

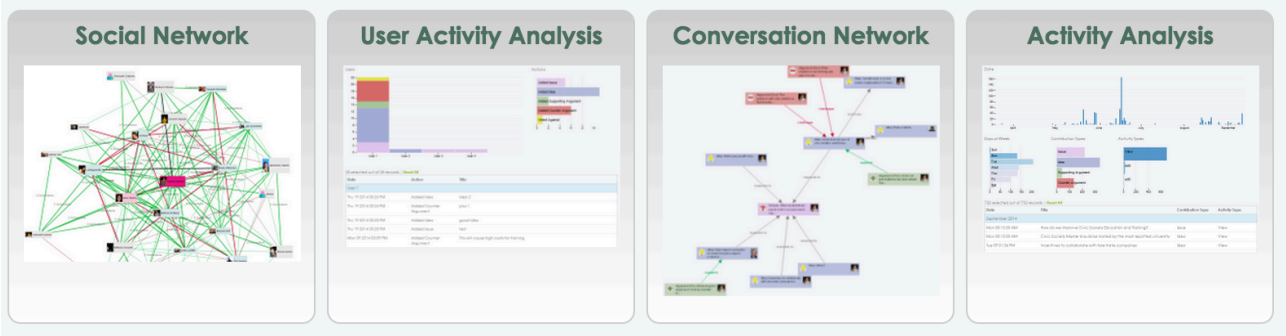

Figure 2: Example of the four tested visualizations

Therefore, each instantiation of the platform delivered feedback through 2 visual devices; we decided to limit the number of devices to 2 to reduce the cognitive burden for subjects deriving from the need to switch from alternative representation. In the combined version (group T4) the subjects had only one type of social visualization and one type of content visualization; in this way their information processing effort has been comparable to those of subject in the other augmented conditions. As we aimed at comparing different platforms, we decided to adopt a between subject experiment design with four different groups. The basic idea behind a between subject experimental approach is that participants can be part of just one group. This type of design is often called an independent measure design because every participant is only subjected to a single treatment. This lowers the chances of participants suffering boredom after a long series of tests or, alternatively, becoming more accomplished through practice and experience, skewing the results. Moreover, this approach has been often used to avoid the carryover effects that can occur in within subjects designs, such as learning, practice and fatigue effects. The problem of this experiment design is that it does not allow to completely controlling the differences among participants. In order to avoid the influences of relevant differences between the groups, they have to be matched or homogenized.

5. Procedure and Data Collection

The test was articulated in four phases:

- *Preparatory phase*: the subjects attended a two-hour seminar about argumentation, with focus on the IBIS formalism and a demo of Catalyst tools (Debate Hub and the Dashboard), and one seminar on the main characteristics the main variables of the crude oil markets. Finally, a warm-up phase of 10 days was organized during which the subjects practiced with the Catalyst tools discussing about a different topic (the future of Greece in the Euro-zone);
- *Experiment phase*: students were requested to discuss through the Debate Hub in a specific time-window (4 hours in one day). The field test was performed on May 21.
- *Administration of a follow up questionnaire* composed of 27 items grouped in cluster. Through the questionnaire we collected data about the subjective

evaluation of the collaborative experience based on the following constructs: perceived quality of collaboration, mutual understanding, and Debate Hub's usability. We also asked subjects to express their individual forecast of what the price trend will be three months after the experiment.

- *Interviews* with a casual sample of participants to collect some direct and informal comment about Debate Hub.

Thanks to the technical features of Debate Hub, we were able to collect and analyse additional data related to the amount of activities and output created by the participants in the discussion (e.g. total number of created ideas, pros, cons, issues, number of votes and so on). Specific auditing features have been added to the DebateHub to allow tracking and recording of what kind of visualization and how many times students used it during the online discussion.

6. Data Analysis and Results

We used SPSS (Statistical Package for the Social Sciences) to undertake the most of our analysis. In general, it is possible to affirm that the treatment groups (that used the augmented platforms - T2, T3, T4) reached better performances than the group which used a plain version of the platform with regard to the i) users' activity level (posts and rating), ii) mutual understanding, iii) perceived quality of online collaboration , iv) perceived ease of use. Instead, there was not the best effect on performance when the two different kinds of visualizations are both provided.
At the start of the online deliberation, an initial map with a framing question and three mutually exclusive answers was presented to the subjects in both groups. In particular, the questions were: What will be the trend of Crude oil price in three months from now? The possible answers were: (i) The price will tend to increase (+10% or higher), (ii) The price will tend to decrease (-10% or lower), and (iii) The price will be stable (+/-10%). The two maps differed solely in the topic (Gold or Oil). When using the platform, participants were required to discuss and map contentious and/or competing points of view in argument maps, with alternative positions, and associated chains of pros and cons.

In the following sections, the results are presented and thoroughly discussed.

6.1 Comparison in terms of Users' activity level

Users' activity is mainly assessed by considering the number of created posts (pros or cons) and the number of expressed ratings. As showed in the table below, the use of augmented platforms, and thus of visualization tools, increase users' activity.

Contribution	T1	T2	T3	T4
#created post	179	196	182	184
#voting	339	430	370	298

Total	518	626	552	482
Average	**14.38**	**17.89**	**15.77**	**14.18**
St. deviation	6.91	14.07	8.15	5.95

Table 2: Descriptive statistics on users' activity

In particular, as it is possible to note, the students of group T2 and T3 were more activity than the other ones. Indeed, the mean scores of these groups are higher respect the others. Basically, these results show that social and content visualization, when provided separately, support effectively users' activity, but, if combined, they are not able to produce the same beneficial effect; in other words, there is not a synergic and interaction effect between social and content visualization. We supposed that these results depend on the information overloading created by the provision of two different visualizations on the users' activity. In general, as the group T4 performed worse than the group T1, which used the plain version, it is possible to suppose that the combination of visualizations creates a diversion, reducing the users' activity. In the Figure 2, we show the overall activity for each groups.

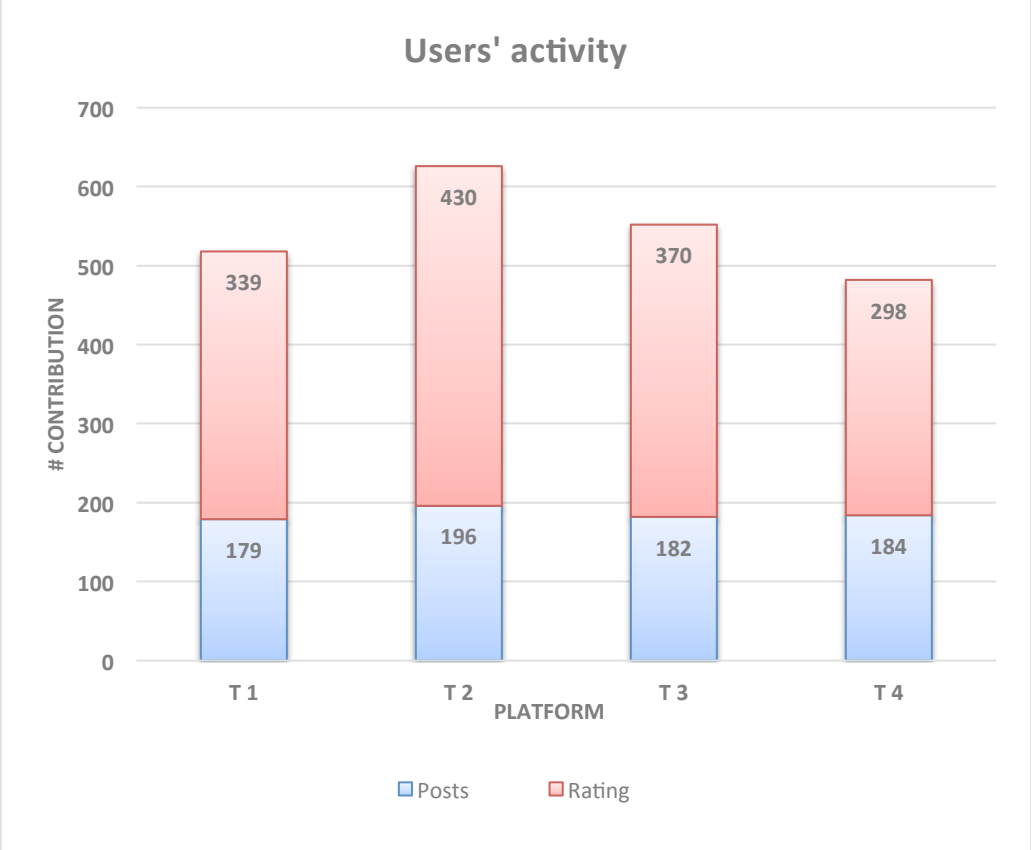

Figure 2: Users' activity per group

As showed in the Figure 3, if we consider the use of the visualization, it is possible to note that the group T4 used most frequently both the visualization devices. We supposed that

the combination of two different typologies of visualizations requires to the students a greater utilization in order that the information could be wholly understood and used to participate in the discussion. On the contrary, the platform T2 and T3, by providing users with related information, this makes easier its understanding and its analysis.

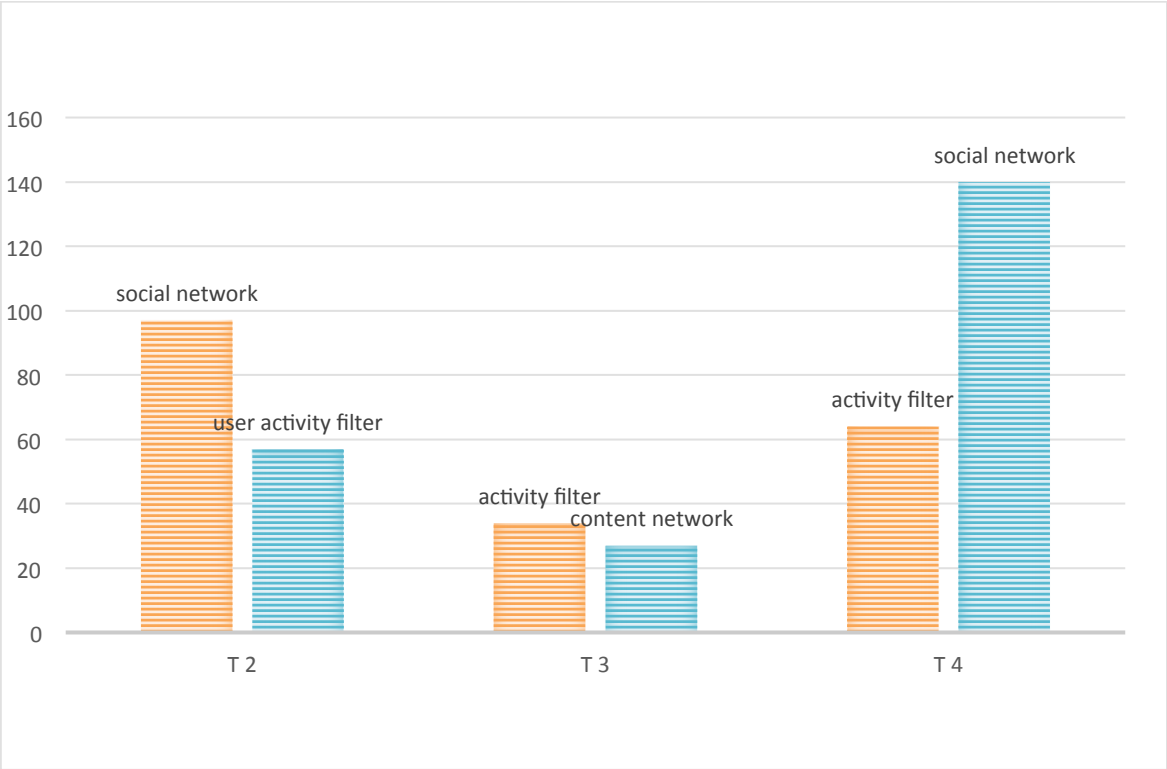

Figure 3: Use of visualization for each group

6.2 Comparison in terms of mutual understanding, quality of collaboration and ease of use of the Debate Hub

Once completed the experiment phase, during which participants discussed about the trend of the future price of the crude oil, a questionnaire composed of 27 items grouped in cluster was administered. Students had to express their opinion on each item by using a 7-point Likert scale. Through the questionnaire we collected data about the subjective evaluation of the collaborative experience based on the following constructs: perceived quality of collaboration, mutual understanding, and Debate Hub's usability. We also asked subjects to express their individual forecast of what the price trend will be three months after the experiment, but we should wait for September for this result. In particular, as showed in the Table 3, the augmented platform performed better than the other one, showing that ***the presence of visualization tools could effectively support the mutual understanding and the collaboration among users; finally, visualization impacts positively also on the ease of use of the platforms.***

	T1	T2	T3	T4
Mutual Understanding	30.08	33.05	31.91	31.61
Perceived Quality of collaboration	42.36	42.88	43.71	42.61
Perceived Ease of Use	27.86	31.23	31.42	30.59

Additionally, as it is possible to note, each visualization impacts on one different users' performance; more specifically, the social visualization supports better the mutual understanding, while the content visualization works well with respect to the perceived quality of collaboration. Finally, it is possible to claim that when there is only one visualization (social or content), the platforms perform better than the other ones with the regard to the perceived ease of use.

6.3 *Results about the students' interviews*

In order to evaluate the strengths and the limitation of the different instantiations of the Debate Hub and better understand the role of diverse visualizations, a set of interviews with a casual sample of participants to collect some direct and informal comment about Debate Hub was performed. In particular, we selected randomly 10 students to interview per group (40 students in total). In general it is possible to claim that students appeared to be "very satisfied about the experience" and they affirmed that "they hope to use again these kinds of platforms also for different tasks and goals". The most part of students suggested using this kind of platform also as an e-learning platform as they could support students to focus on the main aspects of different topics. Several students (about 60%) claimed that "these platforms performed better than forums and social networks in terms of support, orientation and organization of the discussion". Additionally, the students that used the augmented instantiations (T2, T3, and T4) highlighted the "usefulness of the visualizations and of the extra-information in supporting the mutual understanding and the quality of collaboration process".

Students claimed that the best characteristics of Debate Hub are:

i) the knowledge organization which makes very easy to follow the discussion and identify the relevant information,
ii) the rating to express their agreement or disagreement about colleagues' contributions,
iii) the use of short title to summarize in clear and concise way the content of the post, and
iv) the use of URL to provide additional content and information.
v) Finally, about the 45% of students claimed that the platforms were very easy to use

On the contrary, the main limitations are:

i) poor social interaction among participants (45%) and
ii) the scarce visibility of the other users (51%).

iii) Other limitations regard mainly some technical problems such as the slow update of the posts and rating which could depend on the number of students that in the same time used the platforms. Indeed, the same students declared that they did not meet the same problem during the warm-up phase.

7. Conclusion and Implication

The field test aimed at evaluating the performance of Debate Hub and some visualization tools in terms of knowledge accumulation and organization as well as users' participation and satisfaction. More specifically we provided users' with some extra information through two classes of visualization, that is *social awareness* and *content visualizations*.
The results show that ***the augmented platforms performed better than the plain version***. In particular, it is possible to claim that the platforms T2 and T3, that is when only one kind of visualization is provided, reached the better performance both in terms of users' activity and users' perception about mutual understanding, quality of collaboration and ease of use. With the regard to the platform T4, it is possible to claim that, although it improves the users' performance respect to the plain version (T1), it performed worse than the T2 and T3 because of an information overloading.
Therefore, the results of this study has relevant implications for designers of collaborative platforms as it could indicate the way to better define the characteristics of these kind of platforms and mange the trade-off between participation and knowledge organization.
In general, ***students were very satisfied and interested about the experience and they claimed that Debate Hub is a very useful platform in supporting and effectively guiding a decision-making process***. ***Several students suggested using this platform as a learning tool or as a new modality for evaluating their learning.***

7.1. Lessons learned on Experimentation Design

Several new things were learned thanks to this field test about how is it best to engage users in these types of experiemnts. In particular, the definition of a time-window (4 hours in one day) during which participants could discuss about the price trend of the Crude Oil has been one of the most important lesson learned. Indeed, ***the definition of a proper time-window to participate to the discussion has impacted positively both on the quality of the content-generated and on the users' participation***; additionally, preliminary results on the content-generated show that a greater quality of the contributions (no redundancy, stronger coherence of the post respect to the topic, very few off-topic posts etc) was produced. Finally, also the participation was very good as, on an initial community of 143 students, 140 completed all the experiment phases (warm-up phase, experiment phase and compiling follow-up questionnaire).