QLectives – Socially Intelligent Systems for Quality
Project no. 231200

Instrument: Large-scale integrating project (IP)
Programme: FP7-ICT
Deliverable D6.1

[Project periodic report]

Submission date: 12.4.2010

Start date of project: 2009-03-01 **Duration: 48 months**

Organisation name of lead contractor for this deliverable: University of Surrey

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level		
PU	Public	**x**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Document information

1.1 Author

Author	Organisation	E-mail
Terhi Nokkala	University of Surrey	t.nokkalasurrey.ac.uk

1.2 Other contributors

Name	Organisation	E-mail
Nigel Gilbert	University of Surrey	n.gilbertsurrey.ac.uk
Rowena Tanagon	University of Surrey	r.tanagonsurrey.ac.uk
QLectives members	QLectives Consortium	QLECTIVES@LIST.SURREY.AC.UK

1.3 Document history

Version#	Date	Change
V0.1	15.2.2010	Starting version, template
V0.5	5.3.2010	Content report finished
V0.9	10.3.2010	Management and financial report finished
V1.0	11.3.2010	Approved version submitted to EU
Amended version	12.4.2010	Amended version submitted to EU

1.4 Document data

Keywords	Report, progress, management
Editor address data	Terhi Nokkala Centre for Research in Social Simulation University of Surrey Guildford, GU2 7XH United Kingdom

1.5 Distribution list

Date	Issue	E-mail
12.4.2010	Consortium members	QLECTIVES@LIST.SURREY.AC.UK
12.4.2010	Project officer Jose Fernandez-Villacanas	Jose.FERNANDEZ-VILLACANAS@ec.europa.eu
12.4.2010	EC archive	INFSO-ICT-231200@ec.europa.eu

QLectives Consortium

This document is part of a research project funded by the ICT Programme of the Commission of the European Communities as grant number ICT-2009-231200.

University of Surrey (Coordinator)
Department of Sociology / Centre for Research in Social Simulation
Guildford GU2 7XH
Surrey
United Kingdom
Contact person: Prof. Nigel Gilbert
E-mail: n.gilbert@surrey.ac.uk

Technical University of Delft
Department of Software Technology
Delft, 2628 CN
Netherlands
Contact Person: Dr Johan Pouwelse
E-mail: j.a.pouwelse@tudelft.nl

ETH Zurich
Chair of Sociology, in particular Modelling and Simulation,
Zurich, CH-8092
Switzerland
Contact person: Prof. Dirk Helbing
E-mail: dhelbing@ethz.ch

University of Szeged
MTA-SZTE Research Group on Artificial Intelligence
Szeged 6720, Hungary
Contact person: Dr Mark Jelasity
E-mail: jelasity@inf.u-szeged.hu

University of Fribourg
Department of Physics
Fribourg 1700
Switzerland
Contact person: Prof. Yi-Cheng Zhang
E-mail: yi-cheng.zhang@unifr.ch

University of Warsaw
Faculty of Psychology
Warsaw 00927, Poland
Contact Person: Prof. Andrzej Nowak
E-mail: nowak@fau.edu

Centre National de la Recherche Scientifique, CNRS
Paris 75006,
France
Contact person : Dr. Camille ROTH
E-mail: camille.roth@polytechnique.edu

Institut für Rundfunktechnik GmbH
Munich 80939
Germany
Contact person: Dr. Christoph Dosch
E-mail: dosch@irt.de

Project periodic report

Grant Agreement number: 231200

Project acronym: QLectives

Project title: QLectives – Socially Intelligent Systems for Quality

Funding Scheme: Future and Emerging Technologies programme FP7-COSI-ICT

Date of latest version of Annex I against which the assessment will be made: 13th October 2008

Periodic report: 1st x 2nd □ 3rd □ 4th □

Period covered: from 1st March 2009 to 28th February 2010

Name, title and organisation of the scientific representative of the project's coordinator: Nigel Gilbert, Professor, University of Surrey

Tel: +44-1483-68 9173

Fax: +44-1483-689551

E-mail: n.gilbert@surrey.ac.uk

Project website address: www.QLectives.eu

Declaration by the scientific representative of the project coordinator

I, as scientific representative of the coordinator5 of this project and in line with the obligations as stated in Article II.2.3 of the Grant Agreement declare that:

The attached periodic report represents an accurate description of the work carried out in this project for this reporting period;

The project (tick as appropriate):

☑ has fully achieved its objectives and technical goals for the period;
□ has achieved most of its objectives and technical goals for the period with relatively minor deviations;
□ has failed to achieve critical objectives and/or is not at all on schedule.

The public website is up to date, if applicable.

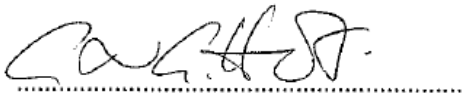
To my best knowledge, the financial statements which are being submitted as part of this report are in line with the actual work carried out and are consistent with the report on the resources used for the project (section 3.6) and if applicable with the certificate on financial statement.

All beneficiaries, in particular non-profit public bodies, secondary and higher education establishments, research organisations and SMEs, have declared to have verified their legal status. Any changes have been reported under section 5 (Project Management) in accordance with Article II.3.f of the Grant Agreement.

Name of scientific representative of the Coordinator: Prof. Nigel Gilbert

Date: ....11...../ ....3...../ ...2010.....

Signature of scientific representative of the Coordinator:



Contents

QLectives – Socially Intelligent ICT Systems for Quality

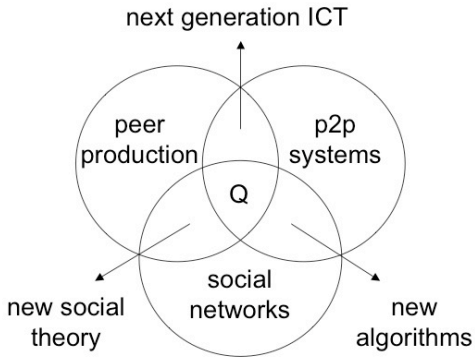
1 Publishable Summary

Introduction

QLectives – Socially Intelligent Systems for Quality – is a multidisciplinary research project funded from FP7-COSI-ICT (Project number 231200). Using a complexity perspective, QLectives will understand, experiment with, design and build cooperative socially intelligent ICT systems composed of self-organising peers, that will enable and support emergent "quality collectives" to enhance, for instance, scientific innovation and decentralised media distribution. The QLectives website can be found at www.QLectives.eu.

Objectives

The QLectives project brings together top social modellers, peer-to-peer engineers and physicists to design and deploy next generation self-organising socially intelligent information systems. The project aims to combine three recent trends within information systems:

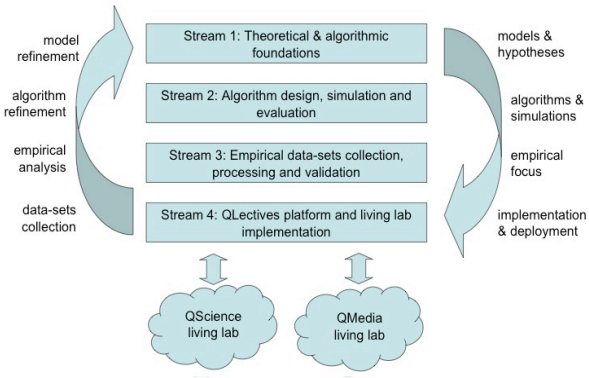


- **Social networks** - in which people link to others over the Internet to gain value and facilitate collaboration
- **Peer production** - in which people collectively produce informational products and experiences without traditional hierarchies or market incentives
- **Peer-to-Peer systems** - in which software clients running on user machines distribute media and other information without a central server or administrative control

QLectives aims to bring these together to form Quality Collectives, i.e. functional decentralised communities that self-organise and self-maintain for the benefit of the people who comprise them. We aim to generate theory at the social level, design algorithms and deploy prototypes targeted towards two application domains:

- **QMedia** - an interactive peer-to-peer media distribution system (including live streaming), providing fully distributed social filtering and recommendation for quality
- **QScience** - a distributed platform for scientists allowing them to locate or form new communities and use transparent reviewing mechanisms to promote quality

The two applications will be based on two existing user communities comprising several thousand people - so-called "Living labs": the media sharing community tribler.org; and the scientific collaboration forum EconoPhysics.



Work performed in Year 1 and significant results

The QLectives project design is based on a constant interaction of empirical data and theoretical models. The project comprises four streams focussing on theoretical and algorithmic foundations; algorithm design, simulation and evaluation; empirical data collection and analysis; and implementation of findings in the QLectives platform and two living labs. These different research steps are cyclically repeated and refined over the four year duration of the project.

Stream 1: Theoretical and algorithmic foundations of quality collectives

In Stream 1 the theoretical foundations of quality collectives have been developed through theoretical and empirical investigations into techno-social systems, trust, social entrepreneurship and scientific collaboration, as well as through modelling.

WP1.1 has investigated the significant literature about models and theories of techno-social systems. This work can be seen as a starting point for QLectives's modelling efforts from two points of view. First, it provides an overview of the existing literature and their main authors, establishing the research scenario from where to start. Second, it integrates the different modelling approaches of the partners who will develop the modelling effort in the project.

In WP1.2 a theoretical framework for a model of the dynamics of trust on networks has been developed. This model, which will be implemented following an agent-based approach, combines processes of trust and social influence. Similarly, initial models of the dynamics of trust on techno-social networks have been implemented and tested in computer simulations. Further theoretical investigations have focused on the various types of conflicts in communities and their impact on trust. The theoretical foundations have been supported by empirical investigations into social entrepreneurship and patterns of scientific collaboration.

In WP1.3 a proposal has been made for a hybrid recommendation algorithm that alleviates the key dilemma of recommender systems: the choice between accuracy and diversity of recommendations. A new similarity index has been implemented for link prediction in complex networks, providing accurate predictions while being computationally less demanding than other standard indices. Algorithms that improve data mining by using time information have been developed. A numerical study of an adaptive recommendation model to assess whether high quality objects can stand evaluation bias and reach their intended goal has been completed. A recommendation algorithm capable of processing gradually arriving information without global re-computation, hence saving computational resources, has been implemented.

The work in WP1.4 has focused on understanding the dynamics of quality collectives using both empirical investigations and agent-based models. The theoretical basis for the model of quality has been developed, and several scenarios describing specific classes of the dynamics have been described. Empirical investigations have focussed on quality in relation to the scientific community.

Stream 2: Algorithm design, simulation and evaluation

In stream 2, a foundation has been laid for translating the models and theories developed in the Stream 1 to algorithms to be implemented in the QLectives living labs (Stream 4).

In WP2.1 several steps have been taken to investigate and design algorithms for emergence of cooperation. We have structured our work into a series of steps. First, areas have been identified for target applications that would benefit from increased cooperation. Second, appropriate user models and possible cooperation-promoting collective mechanisms have been selected. Third, protocols have been designed which apply the user models and collective mechanism models in a way addressing the requirements identified in step 1.

The initial stages of the work on WP2.2 comprise understanding complex situations where users differ in their ability to recognize the quality of objects and where therefore the opinions of some is inherently more valuable than the opinions of others. In some cases, the relevance of objects decays considerably with time. Initial findings show that aggregate quality estimates can be improved if the time when opinions were given is taken into account.

In WP2.3 we studied distributed algorithms for recommendation and ranking (suitable for later deployment in the living labs). Research has focussed on exploring user-based collaborative filtering (CF) methods, which are ideally suited for P2P deployment. Further work is in progress to create databases for training recommender systems based on traces of P2P filesharing systems such as filelist.org.

Stream 3: Empirical datasets collection, processing and validation

In Stream 3 the first batch of data has been collected, standardised according to a unified format created in WP 3.2 and set up in a Data Archive accessible through the QLectives wiki. The first steps in hypothesis testing with the data have also been completed.

In WP3.1, data collection and experimentation, the initial version of the Data Archive containing the "batch 0" databases has been created. The Archive will be used to store all data used in the project and is accessible from the QLectives website. Other data sets have been found and used by members of QLectives. These include Pretty-Good-Privacy algorithm contact graphs, email contact graphs and epistemic hypergraphs**.** The partners managing the two platforms that will become QLectives's living-labs (Tribler and the Econophysics Forum) have implemented mechanisms to collect data about their users' behaviour.

In WP3.2, data processing and knowledge exploration, a unified format has been developed into which all data collected in WP3.1 can be converted. Additionally, the Filelist.org dataset has been converted to unified format and is available from the QLectives wiki.

In WP3.3, the data collected and processed in WP 3.1. and WP3.2 have been used to test hypotheses drawn from the theories developed in WP1.1 and the model and algorithms developed in WP2.2 and WP2.3. Three datasets (drawn from wikis, Flickr and epistemic hypergraphs) have been used for validating hypotheses on models of structure and dynamics of quality collectives.

Stream 4: QLectives platform and living lab implementation

In Stream 4, the QLectives P2P platform, and the updated versions of the two living labs, QScience and QMedia have been implemented. A generic metadata model to represent the quality of any kind of digital object has been created.

WP4.1 delivered version 1.0 of the QLectives Platform, with an initial set of re-usable modules that are the base for the implementation of QMedia and may form the core of a future decentralised version of QScience. The QLectives Platform is built on top of the already deployed and mature P2P tribler.org code-base, which provides most of the low-level P2P functionalities for the social networking and quality facilitation required.

In WP4.2, the new version of the QScience living lab, the Econophysics Forum (http://www.unifr.ch/econophysics) has been launched. The new version is more open to input from its users than the old version, which dated back to 1998, and employs simple recommendation and community-building tools. The Forum includes new functions such as user accounts and personal pages, direct user contributions to all sections of the website except for editorials, and the generation of statistics and trends from user behaviour data.

WP 4.3 delivered version 1.0 of the QMedia platform, addressing an initial set of the required functionalities for this living lab, namely a new and user-friendly graphical user interface, a micropublishing service, and proofs of concept widgets that use the infrastructure and user interface.

WP 4.4 developed a generic metadata model in UML, representing a platform- and implementation-independent model/scheme in terms of software design or software architecture, expressed as class schemes. This will form the basis of developing appropriate algorithms and enables QLectives to use existing metadata types,(e.g the media-optimised TV-Anytime standard), as well as possibly entirely new types of metadata yet to be defined, without the need for defining another complete scheme.

Expected final results and their potential impact and use

The QLectives project has a significant impact on the research into collaboration and P2P activities. Our results will be applied to create two examples of how ICT moulds and becomes part of the systems to which it is applied: QScience – a peer-to-peer application for facilitating scientific innovation by supporting scientific communities, rating activities for quality to identify potential collaborators, hot spots and breakthroughs, and disseminating the right information to the right peers promptly; and QMedia – a peer-to-peer application for transforming media distribution by dynamically identifying shared interest communities and recommending quality contents to them using streaming media technology. We anticipate an impact on all fields in which the collective quality rating of contents and raters can counter an otherwise unsustainable growth in the digital information age. Potential future long-term impacts are expected in the following areas:

- *Next generation social networking:* QMedia currently has only a prototype system for distributed social networking that is not yet deployable. It is on-going work to deliver a social networking function that is both decentralised, efficient and offers privacy options.

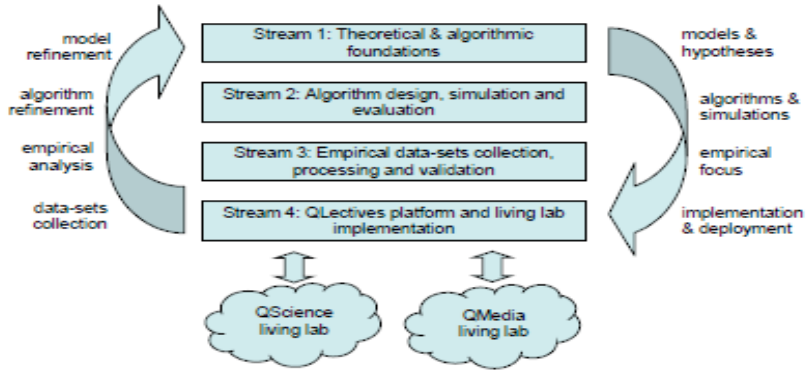
- *Techno-social inclusion:* QMedia allows users to create channels and inject new content (torrents) that can be viewed and searched by other users, without the need for centralised websites or access to other accounts. Currently however, content needs to be indexed by centralised bit-torrent trackers. However, an initial prototype of decentralised tracker functions has been developed for deployment within Tribler. Future releases of QMedia will make use of these functions to provide fully distributed content indexing.
- *Techno-social operating system:* QMedia currently has a prototype system for distributed deployment of P2P-Widgets, which allow for the dynamic injection of code, adding new functionality, into the system. However the currently security model is too limited to allow for general deployment. It is on-going work to address this with possible, redesigned, widget architecture.
- *Scientific collaboration:* Ambitious plans for a new form of scientific collaboration and publication, based on the QScience platform and incorporating the basic functions of an academic journal, a pre-print archive, a bookmarking site, and a social networking site, are being refined.
- The project also supports EU leadership in emerging areas of economic significance by addressing the skills and training gap in P2P technology, and contributing to the building of the EU complexity community.

2 Project objectives for the period

According to the QLectives Work plan, the following objectives should be addressed during Year 1 of the QLectives project. The associated Milestones to be achieved during Year 1 are listed under the relevant objectives.

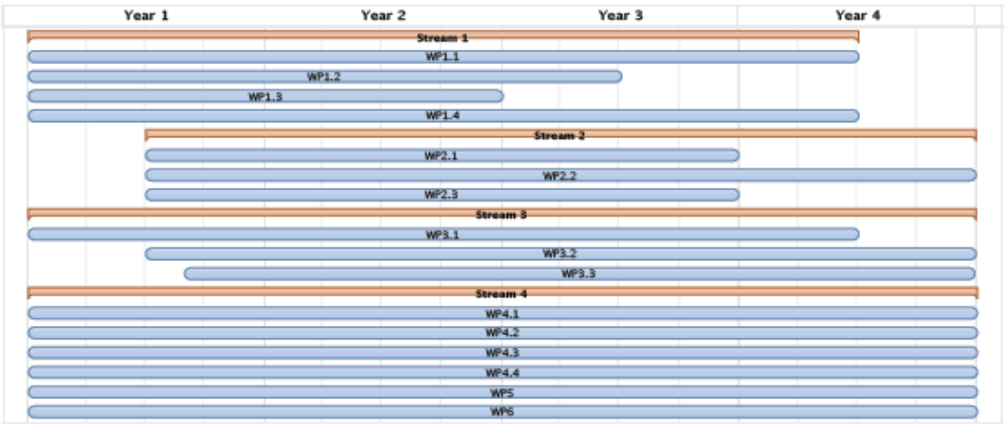
- Understand, experiment with, design and build cooperative socially intelligent ICT systems composed of self-organising peers, that will enable and support emergent “quality collectives” (QLectives).
- Develop theoretical foundations informed by complex system science, to understand the emergence of cooperation and beneficial social structures from interacting peers as the basis of QLectives models. Milestone: M3
- Simulate models to show that they are robust in the face of diverse human behaviour and technological failure. Milestone: M3
- Study simplified models to derive analytically a theory of the macro-level behaviour of such systems. Milestone: M3
- Develop methods for the translation of the models into deployable algorithms that are effective at large scales and that have predictable macro properties. Milestones: M3
- Deploy two living lab ICT social application platforms (QScience and QMedia).
- Collect massive data-sets from the QScience and QMedia living labs (and elsewhere), tracing system behaviour and dynamics. Milestones: M1, M2, M3,
- Apply a scientific approach through rigorous empirical validation, exploration and analysis of datasets, refining and reformulating models and theories. Revise living lab implementations in light of findings. Repeat this cycle throughout the lifetime of the project. Milestones: M1, M2, M3,
- Maintain an open web-based living archive, of datasets collected, freely available to other researchers. Milestone: M1
- Promote an active technology sharing policy by making most of the code available as Open Source.

In order to address the above objectives, the QLectives project is structured into four research streams – characterising it as an integrated project. Each comprises a set of related Work Packages reflecting a particular research approach. The streams support each other but also operate in parallel. The following figure shows a graphical representation of the streams and how they relate to each other.



3 Work progress and achievements during the period

Following the work plan, most work packages have started in Month 1; WPs 2.1, 2.2, 2.3 and 3.2 have started in Month 6, and WP 3.3 in Month 8. The duration of each work package is presented in the following chart.



The month 12 review meeting with the European Commission took place on 26th March 2010. At the request of the reviewers, some amendments were made to the annual progress report after the review meeting. These changes included correcting some of the figures which were incorrect in the earlier version (due to the extremely short reporting period), as well as adding the contributing partners and publications resulting from each of the work package.

Stream 1: Theoretical and algorithmic foundations

Stream 1 produces models, theories, and simulation algorithms based on existing and novel complexity approaches that capture both human agent behaviour, and social and collective structures. Specifically we focus on the emergence of trust and cooperative behaviour and the emergent social structures that incentivise such behaviour. This task is being developed in two phases. First, a class of models and simulations are produced, related to the common functions required in the living lab applications (QScience and QMedia): community evaluation and recommendations for assessing information content quality. Second, hypotheses from our models are elaborated to be empirically tested against our living lab and other collected datasets. Models will be adapted to meet the empirical evidence.

Up to the present point, most efforts in Stream 1 have focused on integrating modelling approaches and building up collaborations among partners. Nevertheless, significant concrete results have already been achieved by the four WPs, in some cases reaching the stage of public dissemination or publication in international fora.

WP1.1 Theory of complex techno-social collectives

Contributing partners

UniS, ETH Zurich, UniFr, UWAR, CNRS

Objectives for WP 1.1

- Develop analytically tractable agent-based models to study the influence of typical characteristics of complex social systems on phenomena such as the emergence, evolution, and decay of social communities
- Extend agent-based models to represent the essential features of complex techno-social systems.
- Derive implications for the improvement of efficiency, stability and resilience of techno-social collectives

Progress towards objectives

The main common achievement of the partners involved in this WP has been the composition of Deliverable D1.1.1 (titled “Overview of theories and models of complex techno-social systems”). This deliverable reviews literature about models and theories of techno-social systems.

Additionally, the QLectives partners have worked separately (individually or by means of bilateral collaborations) in the development of different modelling approaches. Additional work has focused on a game-theoretical approach to study social norms and social phenomena involving cooperation or conflict. It is worth noticing that results in this line (together with D1.1.1 content) are currently being used in Stream 2 as a basis for the definition of new protocols (see deliverable 2.1.1).

Significant results

The deliverable D.1.1.1 can be seen as a starting point for QLectives' modelling efforts from two points of view. First, it provides an overview of the existing literature and their main authors, establishing the research scenario from where to start. Second, it integrates the different modelling approaches of the partners who will develop the modelling effort in the project, so each one of them can have a global overview and realize to what extent they match together and can collaborate.

Publications

Taramasco, C.; Cointet, J. & Roth, C. Academic team formation as evolving hypergraphs. "Paper submitted to Scientometrics." (to appear in 2010). Note: also related to WPs 3.1, 3.2, 3.3.

Helbing, D. & Lozano, S. "Routes to Cooperation and Equilibrium Creation in the Prisoner's Dilemma". Submitted to Physical Review Letters.

Helbing, D.; Szolnoki, A.; Perc, M. & Szabó, G. "Evolutionary establishment of moral and double moral standards through spatial interactions". Submitted to *PLoS Computational Biology*.

Tabourier, L.; Roth, C. & Cointet, J. Generating constrained random graphs using multiple edge switches. Paper submitted to ACM JEA Journal of Experimental Algorithmics.

Cointet, J. & Roth, C. (2010). Local circles, local topics: structural and semantic proximity in blog networks. Paper submitted for review for 4th Intl AAAI Conf on Weblogs and Social Media, Washington, DC, 23-26 May 2010.

Helbing, D.& Nowak, A. (2009). Dynamika Tłumu / Dynamic of the Crowd. In: Wasilkowska, A.& Nowak, A. (eds.) (2009). Warszawa jako struktura emergentna: Em_Wwa 1.0; Warsaw as Emergent Structure: Em_Wwa 1.0 Warszawa. Warsaw: Fundacja Bęc Zmiana, pp. 55-61.

Helbing, D. (2009). The Role of Mobility for the Outbreak of Cooperation among Egoists, Meeting of the Leopoldina, German Academy of Sciences, Zurich, Switzerland, 5 March 2009.

Helbing, D. (2009). Spontaneous Outbreak and Breakdown of Human Cooperation, Minerva International Workshop on "The Science of Complexity", Eilat, Israel, 30 March 2009.

Helbing, D. (2009). Behavior of crowds and socio-inspired technology: prospects for a dialogue between the engineering, natural and social sciences, series of events on simulation and data analysis IAS (Institute for angewa), Wädenswill, Switzerland, 8 May 2009.

Helbing, D. (2009). How Social and Collective Human Behavior are Born from Simple Individual Interactions, Meeting of the Cognitive Science Society, Symposium: The Emergence of Collective Structures Through Individual Interactions", Amsterdam, Netherlands, 31 July 2009.

Helbing, D. (2009). Cooperation, Norms, and Revolutions: A Unified Game-Theoretical Approach, Keynote Talk at ECCS 2009, Warwick, United Kingdom, 21 September 2009.

Helbing, D. (2009). Towards Probing Foundations of Society by Computer Simulations, Max-Planck Institute for Evolutionary Biology, Plön, Germany, 13 November 2009.

Helbing, D. (2009). Cooperation, Norms, and Conflict: Towards Simulating the Foundations of Society (Keynote Talk), EURACE Workshop, Genoa, Italy, 20 November 2009.

Helbing, D. (2009). Cooperation, Norms, and Conflict: Towards Simulating the Foundations of Society, Conference on Rational-Choice: Theoretical models and empirical applications, Venice, Italy, 3 December 2009.

Helbing, D. (2009). Cooperation, Norms, and Conflict: Towards Simulating the Foundations of Society, Paris Interdisciplinary PhD Symposium "From sparse entities to crowded environments: Numbers in living systems", Paris, France, 7 December 2009.

Roth, C. (2009). Co-évolution entre réseaux sociaux et opinions: le cas de la blogosphère. Invited talk at the MASHS 2009 conference in Toulouse, France, 8-9 June 2009.

Roth, C. (2009). Reciprocal influence of social and semantic patterns in a dynamic blog network. Invited communication at the DOOCN- III workshop at ECCS 2009, Warwick, UK, 23-24 September 2009.

D1.1.1 Overview of theories and models of complex techno-social systems

WP1.2 Complexity models of trust networks

Contributing partners

UniS, ETH Zurich, UniFr, UWAR

Objectives for WP1.2

- Review, compare and identify shortcomings of existing complexity models of agency for techno-social systems
- Develop algorithmically specified novel agency models tuned for techno-social communities
- Develop “informational division of labour” interaction scenarios and games
- Apply and tune psychologically plausible “dynamic social impact theory” for techno-social communities

Progress towards objectives

One of the main goals of this WP is to “Apply and tune psychologically plausible ‘dynamic social impact theory’ for techno-social communities”. Four actions have been taken in this direction.

First, a theoretical framework for a model of the dynamics of trust on networks has been developed. This model, which will be implemented following an agent-based approach, combines processes of trust and social influence. Conflict in groups has been identified as one of the main factors damaging trust, potentially affecting their capability to produce quality products and services.

Second, strategies of social entrepreneurs were analysed. It was revealed that social entrepreneurs achieve social change mainly by building networks of trust in the social groups they work for. They do it by a sequence of actions where each action is based on the success of the previous action. Based on the results of the research a proposal for a book was prepared, which was accepted by Cambridge University Press: Praszkier R, Nowak A, (in preparation) *Social Entrepreneurship: Theory*

and Practice, Cambridge University Press. Research concentrating on changes in social groups where social entrepreneurs work has been initiated.

Third, initial models of the dynamics of trust in techno-social networks were implemented and tested in computer simulations. The models of the development of trust were integrated with the dynamical theory of social impact. In these models individuals interact on networks to assess how much trust can be given to each of the interaction partners. The amount of trust is a component of "strength" which is one of the main variables in the dynamical theory of social impact. Computer simulations have compared the accuracy of judgments of a model with trust updating and one in which there was no trust. The results revealed that models with trust built into decision mechanisms function better than the models without this mechanism.

Finally, research has been conducted to analyse academic collaboration. This is based upon a survey of 249 researchers combined with interview data from 13 respondents, integrated with data collection for WP1.4. So far analysis of the interview data highlights the importance of social connections and physical interactions for establishing trusting relationships, and subsequently, collaborations. In particular, the survey data reveals that although a number of Web 2.0 social tools are available to enable networking online, these currently do not seem to be able to offer a proxy for traditional methods of establishing trust. Additionally, disciplinary differences in perception of collaborative practices and the intensity and quality of those relationships have been noted.

Significant results

A review of the related literature has revealed that with respect to the impact on trust two types of conflict can be distinguished: constructive conflicts and destructive conflicts. Destructive conflicts are aimed at hurting the other side and damage trust and the quality of group products. Constructive conflicts, in contrast, are aimed at resolving issues, increase trust and have beneficial effects for quality. Two studies of Internet conflict were conducted by the consortium. The results revealed that the constructiveness and destructiveness of conflicts represent two independent (orthogonal) dimensions. A draft of an article was written (Samson, K, Nowak A., 'The Orthogonality of Conflict Processes: On Being Destructive and Constructive at the Same Time').

The Open Source movement has been identified as a techno-social network of high quality. The focus of study was on conflicts in open source communities. Initial qualitative analysis revealed the existence of two typical types of conflict. Ideological conflict is usually initiated by newcomers to the group. Examples are attitudes towards some commercial software, or whether to make the code publicly available. Such destructive conflicts have a tendency to last longer than other types of conflict and in some instances they are associated with dissatisfaction and possible damage to trust. The second type of conflict is centred on low level technical issues. These conflicts are resolved quickly and result in common satisfaction, and perhaps an increase in trust. These conflicts likely to belong to the constructive type.

Publications

Helbing, D. & Johansson, A. (2010). "Evolutionary Dynamics of Populations with Conflicting Interactions: Classification and Analytical Treatment Considering Asymmetry and Power". *Physical Review E* 81, 016112, 2010.

Samson, K, Nowak A., 'The Orthogonality of Conflict Processes: On Being Destructive and Constructive at the Same Time.

Helbing, D. (2009). Cooperation and Conflict in the Prisoner's Dilemma and the Emergence of Norms, CCSS Workshop in ETH Zurich: Coping with Crises in Complex Socio-Economic Systems, Zurich, Switzerland, 8 June 2009.

Helbing, D. (2009). Cooperation and Conflict in the Prisoner's Dilemma and the Emergence of Norms, Workshop on Traffic: Managing the Complexity of Large Logistic Transport Systems by bio-inspired strategies, Venice, Italy, 3 July 2009.

Helbing, D. (2009). Cooperation and Conflict in the Prisoner's Dilemma and the Emergence of Norms, Autumn Meeting of the Modeling and Simulation Section of the German Society for Sociology: "Social Networks and Social Integration", Groningen, Netherlands, 24 October 2009.

Helbing, D. (2009). Self-Organization and emergence in social systems. Modeling the co-evolution of social environments and cooperative behavior, ATACD Conference: "Changing cultures: Cultures of change", Barcelona, Spain, 10 December 2009.

WP1.3 Theory of quality emergence in complex systems

Contributing partners

UniS, ETH Zurich, UniFr, UWAR

Objectives of WP 1.3

- Developing agent-based models for self-consistent estimates of quality
- Proposing and investigating effective approaches to quality signaling
- Deriving implications for quality signaling in real applications.

Progress towards objectives

Several results have been achieved towards these objectives. A proposal has been made for a hybrid recommendation algorithm that alleviates the key dilemma of recommender systems: the choice between accuracy and diversity of recommendations. A new similarity index has been implemented for link prediction in complex networks. This similarity index provides very accurate predictions and is computationally less demanding than other standard indices. Algorithms that make use of time information (e.g., the exact time when a rating was given by a user) to improve data mining have been developed.

Additionally, a numerical study of an adaptive recommendation model has been completed, aiming to assess whether high quality objects can stand evaluation bias and reach their intended audience (evaluation bias includes both intrinsically imperfect evaluations as well as the evaluations of intentionally malicious users). A recommendation algorithm has been implemented, capable of processing gradually arriving information without the need for global re-computation and which hence saves computational resources.

The steps taken also include collecting data on scientists' publishing patterns and attitudes to 'open science'. This is based upon a survey of 249 researchers combined with interview data from 13 respondents, integrated with data collection for WP1.4. Traditional citation data with usage and community metrics are also being investigated.

Analysis of interview and survey data of attitudes and practices in relation to 'open science' reveal interesting discrepancies and conflicts. Whilst most scientists agree that 'open science' is a good thing, and that ultimately it will lead to a better quality of research and more scientific discoveries, this is but one aspect of a larger picture of scientific practice. Revealing links with WP1.2, it may be noted that trust is a very important factor for scientists intending to share their results or methods, and is predominantly related to the scientific culture of primacy. Institutional and governmental bodies may both encourage and discourage open scientific practice. Investigations of citations, usage and community metrics indicate that there is not always clear equivalence of measures and that often a single measure (e.g., impact factor) does not in reality clearly measure one individual phenomenon.

Further work undertaken by the QLectives partners is related to developing a model about the emergence of standards of quality, and group dynamics in the emergence of quality judgements i.e. how quality judgements emerge in a group discussion. This is similar to social influence, a process with a clear product: standards of quality and eventually ranking of objects with respect to their quality.

Finally, empirical studies on the emergence of standards of quality have been conducted. In the initial empirical study, groups of several individuals were instructed to rank order a set of pictures with respect to quality. Three groups of students were asked to rate a set of pictures. Their interaction has been recorded on video and the conversation was transcribed in preparation for a full analysis.

Significant results

A paper reporting the hybrid recommendation algorithm mentioned above has been accepted for publication in *Proceedings of the National Academy of Sciences* (PNAS). This achievement makes the work developed within QLectives highly visible on a world-wide scale. A report on the data about scientific collaboration has been accepted for poster presentation at the Fourth International AAAI Conference on Weblogs and Social Media (ICWSM 2010) in Washington DC.

Publications

Zhou, T., Kuscsik, Z., Liu, J.-G., Medo, M., Wakeling, J. R. and Zhang, Y.-C. (2010) Solving the apparent diversity-accuracy dilemma of recommender systems. PNAS, 107 (10).

Jiang, L-L., Medo, M. Wakeling J.R, Zhang, Y-C.& Zhou, T. Building reputation systems for better ranking. http://arxiv.org/abs/1001.2186

Helbing, D. (2009). A kinetic approach to game theory: From opinions to norms, Workshop: Kinetic and Mean-field models in the Socio-Economic Sciences, Edinburgh, United Kingdom, 27 July 2009.

Scissors, L.; Gill, A. J., & Gergle, D. (2009). "You Can Trust Me", "I Can Trust You": Linguistic Accommodation and Trust in Text-based CMC. Paper presented at the 2009 Annual Conference of the International Communication Association, Chicago, IL, May 2009.

D1.3.1 Simulation of quality emergence

WP1.4 Modelling the dynamics of quality collectives (UniS)

Contributing partners

UniS, ETH Zurich, UWAR, CNRS

Objectives of WP1.4

- Understand the dynamics of quality collectives using agent-based models: how they emerge, consolidate and decline
- Create an abstract model of the dynamics of quality collectives
- Apply this model to understanding the dynamics of quality collectives observed in the QScience and QMedia living labs

Progress towards objectives

To date work has focused on addressing the first objective of WP1.4 (to 'Understand the dynamics of quality collectives using agent-based models: how they emerge, consolidate and decline'), and working towards deliverable D1.4.1 'Modelling the dynamics of quality collectives'. The theoretical basis for the model of quality has been developed and several scenarios describing specific classes of the dynamics have been described. The model has been presented at the Quality Commons workshop (Andrzej Nowak (2010) "The Social Dynamics of Quality Choice", Paris, 28th-29th January 2010) and an article is currently being prepared.

Further research has examined the role of scientific communities in establishing quality, and in particular, the role of online tools and collaboration facilities for encouraging the production of high quality science. A survey of 249 researchers combined with interview data from 13 respondents, and a functional and discourse analysis of relevant online tools has led to a submitted conference paper (Gill, A., Nokkala, T. and Gilbert, N., Patterns of Adoption of Science 2.0. Paper submitted to the Fourth International AAAI Conference on Weblogs and Social Media); a conference paper (Gill, A. and Gilbert, N.: Web 2.0, Its Influence in Academic Software, and the Future of Academic Collaboration. Paper presented at the European Computer Supported Cooperative Work Conference (ECSCW2009) Workshop, Academia 2.0 and Beyond. Vienna, Austria, September 2009.); and a journal manuscript in preparation (Gill, A., Nokkala, T. and Gilbert, N. Web 2.0 tools in the Science 2.0 workplace: Functionality and the presentation of problem-solution).

Significant results

High quality is associated with the belief that it cannot be popular (popular=cheap). These dynamics share important elements with the minority game, and are combined with the majority game. Individuals want to be in the groups that established new quality standards, but also want to be elite. Therefore the attractiveness of this group is somehow (perhaps nonlinearly) related to its size: as the group grows its memberships loses attractiveness. The game that people appear to be playing is to start a new trend, but at the same time to be in a small majority. This dynamics describes the dynamics of fashions. One just needs to be the leader in the quality game, i.e. have objects that are at the top of quality standards. In cases where an individual cannot compete, they

try to set up new standards of quality that would allow them to assume dominance. An individual incapable of winning a game may attempt to change the rules of the game (i.e. the measure of quality). On the other hand, the dynamics of quality relating to products not associated with prestige (i.e. washing powder or light bulbs) appear mostly concerned with converging in a majority (social influence) game.

Findings of the abovementioned research show that Web 2.0 technologies can potentially change the way that science is performed: It offers new possibilities for scientists to engage in the design and creation of tools and resources useful to them, in addition to greater openness and social connection. Emerging patterns of tool use and collaboration indicate that not all scientists are adopting these technologies equally, with age and academic position having an effect. However, more interestingly, disciplinary affiliation often determines scientists' needs and perceptions of technology. While those working in the Human Sciences often reported using Web 2.0 technologies during the course of their research, this was often as an object of study. Conversely, those working in the Natural Sciences were more likely to report using Web 2.0 tools to support collaboration in collocated research groups or of a distributed nature. Discourse and functional analysis of selected Web 2.0 tools shows that they often present themselves as agents to assist the democratising of access to technology and resources, but are also sure to emphasis their professional and reliable credentials. In terms of tool functionality, this could be broadly separated into social, search, compatibility and document management aspects, with search and compatibility particularly targeting academic users.

Publications

Brax, N. & Amblard, F. (2009). A self-repairing solution for the resilience of networks to attacks and failures. Presented at the European Conference on Complex Systems 2009, University of Warwick, UK, 21-25 September 2009

Gill, A. J.; Nokkala, T. & Gilbert, G. N. (2010). Patterns of Adoption of Science 2.0. Paper submitted for review for 4th Int'l AAAI Conference on Weblogs and Social Media, George Washington University, Washington, DC, 23-26 May 2010.

Gill, A. J. & Gilbert, G. N. (2009). Web 2.0, its influence in academic software, and the future of academic collaboration. Paper presented at the Academia 2.0 and beyond – How Social Software changes research and education in academia workshop at the European Conference on Computer-Supported Cooperative Work, Vienna, Austria, 8 September 2009.

Lozano, S. (2009). Techno-social systems supporting socially constructive processes in a self-organized fashion, European Commission's Expert Consultation Workshop "Shaping the Future" , Brussels, Belgium, 4 November 2009.

Stream 2: Algorithm design, simulation and evaluation

Based on the high-level models, theories, and simulations developed in Stream 1, Stream 2 translates these to algorithm designs that can be implemented and deployed in our living labs. The resulting deployable algorithms need to factor in practical aspects such as the efficient storage/recall of data and other engineering constraints. This stream focuses on two distinct ways of importing insights from Stream 1 models into algorithms. Firstly, by using the models of human agent motivations it is possible to test (in simulation) the effectiveness of the proposed QLectives platform algorithms. Secondly, by identifying desirable emergent properties, it is possible to base algorithms and protocols directly on complex systems models, even when these may have been designed to model human systems.

While all three work packages included in Stream 2 only started in month 6 and many of results achieved within Stream 1 have become mature and available only recently, the progress of Stream 2 is considerable. Progress in each work package is described in the following.

WP2.1 Algorithms for the emergence of cooperation

Contributing partners

TUD, ETH Zurich, USZ, UniFr

Objectives for WP2.1

- Translate novel theoretical models of cooperation formation into algorithms for ICT
- Simulate the performance of these algorithms under realistic assumptions based on target applications
- Incorporate human agent behaviour models
- Propose algorithms for implementation and deployment in Stream 4

Progress towards objectives

In order to achieve the planned ends, work was structured into a series of steps:

1. Identify *areas for target applications* that would benefit from increased cooperation.
2. Select *appropriate user models* which may capture the behaviour of users within such application areas.
3. Determine the *possible collective mechanisms* that can promote cooperation within the application areas for the given user models.
4. Design *appropriate protocols*, that could be deployed, that instantiate the collective mechanisms.

Step 1 involves determining high-level requirements for given application areas that require high levels of cooperation between peers. In step 2 user behaviour models are selected, informed by theoretical models (such as those discussed in deliverable D1.1.1) but plausible in the context of actual user behaviour in P2P systems. Step 3 draws on existing theoretical models which propose collective mechanisms which may be applicable to the application areas we have selected and the user models we will consider. Step 4 is perhaps the most difficult step since it requires the design of protocols that apply the user models and collective mechanism models in a way that addresses the requirements identified in step 1. For this reason step 4 comprises two parts. The first is the use of simulation to test potential ideas for protocol designs and the second, the implementation of potential protocols within stream 4 of the QLectives project (comprising WP4.1-WP4.3).

Step 1

The following two requirements for the QMedia platform were identified:

1. To produce the levels of cooperation (seeding) found in private communities without the need for a central website or administration. That is, we wish to provide a distributed incentive system of sufficient quality to support efficient downloading and video-on-demand.
2. To provide tools and incentives promoting quality user contributions in the form of rating, moderation and spam prevention without the need for a centralised website or administration. The aim is hence to produce a self-managing and self-policing system sufficient to support high levels of community quality.

These requirements should lead to efficient media sharing, high performance download, and quality contributions from community members in QMedia.

Step 2

In the next step, several broad user model variants were identified. They can be found in a number of theory models related to the emergence of cooperation (and other phenomena):

1. *Default model*: It is often found that users within P2P systems appear to be doing little more than running their client software for some time without any significant interaction.
2. *Rational model*: It indicates a user that will pursue their own self-interest using the controls they have over the protocol (Shneidman and Parkes, 2003). (Here we discount the users hacking or changing the client protocol.)
3. *Altruistic model*: Although rarely seen in theoretical models, several measurement studies of P2P file-sharing communities provide evidence of users who appear to behave in altruistic ways. For example, seeding behaviour in public communities and the maintenance and administration of central websites for private communities.
4. *Evolutionary model*: Evolutionary models of behaviour have been widely employed within evolutionary game theory and agent-based modelling (Axelrod, 1984). The assumption is that behaviour variants (or strategies) are copied between agents based on some measure of success (or fitness), often termed utility. The assumption is that agents can determine how well they are performing using some metric and can also determine how well others are doing by the same metric. Agents can then copy (imitate) the strategies from those who outperform them.
5. *Satisficing model*: Satisficing models (Simon, 1955) assume that agents are not attempting to maximise a utility but rather have some aspiration threshold that they will be satisfied with. This model requires fewer assumptions than both a rational model and an evolutionary model. It only requires that an agent can compare its own performance against its aspiration threshold.

Step 3

The following collective mechanisms have been proposed in theoretical models and experiments:

1. *Indirect reciprocity*: Indirect reciprocity mechanisms allow cooperative interactions to occur between strangers who have never met before and may never meet again (Nowak and Sigmund, 2005).
2. *Migration and group selection*: Migration models allow agents to move or "migrate" (change their interaction neighbours) based on some local performance criteria (Helbing and Yu, 2009). Migration mechanisms are important within the wider set of models known as group selection models (Traulsen and Nowak, 2006).
3. *Altruistic punishment*: Experimental evidence (Fehr and Gintis, 2007) shows that adding the possibility of altruistically punishing defectors to public-goods games significantly increases cooperation levels. An altruistic punishment happens when an agent decides to incur some cost to reduce the benefit a non-cooperator agent receives in the game.

These can be seen as potential templates for distributed protocols that, under given user models, could self-organise the system to high levels of cooperation and other quality outcomes.

Step 4

Many private peer-to-peer file sharing communities implement credit policies to incentivise users to contribute upload resources. Results of the agent-based simulation show that credit policies, based on bandwidth contribution and satisficing, altruistic and default user models, can lead to both "crunches" and "crashes" where the system seizes completely due to too little credit or too much credit. The conditions leading to these system pathologies were explored and a theoretical analysis presented, allowing determining whether a community is sustainable or whether it will eventually crunch or crash. The analysis was also applied to produce a novel adaptive credit system that automatically adjusts credit policies to maintain system sustainability (under the assumption of a satisficing model).

Significant results

Based on the preparatory work described in steps 1-3, extensive agent-based simulations were performed in step 4, leading to a number of interesting results. Firstly, if there is too little or too much credit in the system (i.e. if the proportion of the rich at the start is ≤ 0.2 or ≥ 0.8) then the final state of the system is a crunch (where all peers are seeding) or a crash (where no peers are seeding). This leads system throughput (amount of data exchanged between peers and hence a measure of system performance) to go to zero. Too much seeding is just as bad as too much leeching (purely downloading) because for seeding to increase system throughput there must be a matching peer who wishes to download. Interestingly, it was also found that even when only a small number of users followed an altruistic user model (in which users contribute far more bandwidth than necessary to support their downloading requirements), this degrades the performance of the system as a whole due to the hogging of credit. This indicates that simple credit systems function poorly even if a small number of users behave altruistically - which is counterintuitive and less than desirable.

Some on-going work has considered alternative credit policies, based on effort rather than contribution, which can ameliorate some of the problems associated with inequality within private communities (Rahman et al, 2010). Such an approach rewards peers with credit based on the amount of effort (what proportion of available resources are contributed) rather than total contribution (what amount of resources are contributed). This means that the user is rewarded for good behaviour rather than just for absolute contribution. Our simulations of an effort based approach under the assumption of a default user model (with both fast and slow upload bandwidth peers) show that under the effort based policy both fast and slow peers perform better than in the contribution based approach. Full details of these simulations can be found in (Rahman et al, 2010).

Publications

Rahman, R.; Meulpolder, M.; Hales, D:, Pouwelse J. & Sips, H. (2010). Improving Efficiency and Fairness in P2P Systems with Effort-Based Incentives. To be published in *Proceedings of the IEEE International Conference on Communications*, Cape Town, South Africa, 23-27 May 2010.

Hales, D.; Rahman, R.; Zhang, B.; Meulpolder M. & Pouwelse J. (2009). BitTorrent or BitCrunch: Evidence of a credit squeeze in BitTorrent? Proceedings of the 5th Collaborative Peer-to-Peer Systems (COPS) Workshop, in conjunction with 18th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises, Groningen, the Netherlands, 29 June - 1 July 2009.

Hales, D., (submitted) Mix, Chain and Replicate: Methodologies for Agent-Based Modelling of Social Systems. In Mollona, E., (ed) *Computational Organisational Theory*. Computer simulation as a new paradigm for research in social sciences, Routledge.

Hales, D., (submitted) Rationality meets the Tribe: Recent Models of Cultural Group Selection. In Mollona, E., (ed) *Computational Organisational Theory*. Computer simulation as a new paradigm for research in social sciences, Routledge.

Gill, A. J. (2009). Social Information in Computer-Mediated Communication. Invited talk at the HCI Research Group Seminar Series, University of York, 29 October 2009.

Hales, D. (2009). BitTorrent or BitCrunch: Evidence of a credit squeeze in BitTorrent? Presentation at the 18th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises - WETICE 2009 (COPS workshop), Groningen, the Netherlands, 29 June 2009.

D2.1.1 Candidate theory models for cooperation algorithms

WP2.2 Algorithms for the emergence of collectives

Contributing partners

UniS, TUD, ETH Zurich, USZ, UniFr, IRT

Objectives for WP2.2

- Develop algorithms for detection and reinforcement of collectives
- Test the performance of the proposed algorithms by simulations
- Provide a basis for efficiently distributed implementation of the proposed algorithms

Progress towards objectives

Work on this work package began only in month 6 and the corresponding deliverable is due in month 48 (D2.2.1). The initial stages of the work on WP2.2 comprise understanding complex situations where users differ in their ability of recognizing high-quality objects and hence their opinion is inherently more valuable than the opinion of others.

Significant results

The research findings show that aggregate quality estimates can be considerably improved if the time when opinions were given is taken into account. This is particularly true in a case when the relevance of all objects decays with time and new objects appear constantly (as for news, for example). The result is a highly dynamic system where objects appear and attract some attention, until their relevance diminishes and users' interest shifts to fresher objects. The research also shows that when a given recommendation model (Medo, Zhang, Zhou, 2009) is generalized by introducing user reputation, spam objects can be entirely eradicated from the system by giving high enough weight to reputation. Even more, an intermediate weight of reputation is both able to suppress spammers effectively and improve recommendation performance.

Publications

Whalen, J. M.; Pexman, P.M., & Gill, A.J. (2009). "Should Be Fun—Not!": Incidence and Marking of Nonliteral Language in E-Mail. Journal of Language and Social Psychology, September 2009; 28: 263 - 280.

Jin, C-H., Liu, J-G., Zhang, Y-C. & Zhou, T. Adaptive information filtering for dynamic recommender systems. http://arxiv.org/abs/0911.4910

Gill, A. J.; Nowson, S. & Oberlander, J.(2009). What are they blogging about? Personality, topic and motivation in blogs. Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media. San Jose, California,17-20 May 2009.

Gill, A. J.; Nowson, S. & Oberlander, J.(2009). What are they blogging about? Personality, topic and motivation in blogs. Paper presented at the 3rd International AAAI Conference on Weblogs and Social Media. San Jose, California,17-20 May 2009.

WP2.3 Algorithms for trusted and quality rankings

Contributing partners

TUD, ETH Zurich, USZ, UniFr, CNRS, IRT

Objectives for WP2.3

- Design algorithms for exploiting partial trust and rating information
- Design fully distributed robust implementations of these algorithms
- Help integrating designed algorithms into the living lab platforms

Progress towards objectives

The main objectives for this period were designing algorithms for recommendation and ranking, and designing distributed version of these algorithms for deployment in the living labs. (Actual deployment in living labs is scheduled to be reported on in month 36 in deliverable D2.3.1).

As a first step, a thorough survey has been carried out on the state of the art of recommender systems, and in particular, distributed and P2P implementations of recommender systems. The most promising approaches for a distributed implementation have been identified. One of the main criteria for evaluating recommender systems is that the communication resource limitations have to be respected at every participating node, as well as being able to achieve a sufficiently fast convergence speed. In related work, there are two main approaches: content based approaches and collaborative filtering approaches. In the latter case, recommendation is content-agnostic, i.e., a set of items of unknown content and a set of users are assumed, and a number of ratings of some of these items by the users is assumed to be known. The task is then to predict the ratings of the unrated items.

Due to their content-independent nature (and therefore generality) as well as their rather good performance, and their focus on networks, we decided to focus on collaborative filtering (CF) in this phase. We identified a number of requirements that a P2P CF algorithms needs to have. One of the main requirements is that at each node there can be communication only with a limited number of other nodes. This eliminated many approaches. For example, singular value decomposition calculates a simplified version of the (user-item) rating matrix thereby extracting features that can be used to calculate user-similarity even if the rating matrix is sparse. However, singular value decomposition is very difficult to calculate in a P2P fully distributed fashion, because it involves iterative methods using non-sparse matrices thereby incurring a huge network load. Another class of methods that are not suitable are the so-called item-based methods, where item similarity needs to be calculated. Since network nodes naturally map to users, and not items, calculating item similarity necessarily involves P2P data structures such as distributed hash tables, and this increases the cost of the methods significantly.

Overall, the choice was made to explore user-based CF methods, because they are ideally suited to a P2P setting, and they provide a competitive quality as well. A set of user-based CF methods involving a number of similarity measures and other parameters were implemented in the PeerSim simulator. A P2P algorithm was proposed for computing the k-nearest neighbour (KNN) overlay network based on any similarity metric. This KNN network makes it possible to implement user-based CF, because there one needs to derive the prediction for a given rating based on the weighted ratings of the k most similar users. A number of baseline algorithms were implemented, such as the BuddyCast algorithm in the Tribler platform. The performance of these alternatives were compared on a number of different databases, according to a number of different performance metrics.

The main performance metrics selected were prediction performance (as usual) and load balancing, which is considered very important for a P2P system. Even though the kNN network appears to be sparse, it was found to have a degree distribution that can be very unbalanced thereby generating huge surges of load on the central nodes, especially in a large system. Algorithms need a remedy for this behaviour if they are to be practically feasible. The research demonstrated that algorithms providing good performance as well as an evenly distributed load could be designed.

It was also realised that characteristics of the database at hand have a large effect on the performance. Sparse rating matrices appear to generate an extremely unbalanced kNN network, for example, while rating matrices with less sparsity have a less unbalanced degree distribution. This observation calls for a closer analysis that is currently being carried out. Additionally, databases are being created for training recommender systems based on traces of P2P filesharing systems such as filelist.org. Preliminary databases have been created, and different options are currently being explored. The aim is to be able to predict user preference based on the behaviour of a filesharing (bittorrent) client, such us download and removal of files, length of seeding, etc.

In parallel with this, modelling of one-way connections in P2P systems with firewalls and NAT devices are being investigated. This seems to be a major source of problems in practical protocols, while little is known about this in the P2P simulation literature. In terms of distributed protocols this is an important topic of research for the QLectives project. Several very large traces have already been pre-processed, and important data related to firewall behaviour has been extracted.

Significant results

Focusing on a fully distributed P2P implementation of existing recommendation algorithms, the QLectives research has shown that load balancing plays a more important role compared to centralized implementations, where this aspect is of no significance. Previous algorithms can be significantly improved so that they respect the resource limits at each node while providing the same or even better prediction performance. In general, the conclusion is that in fully distributed implementations, resource limitations (especially those related to communication) play an important part that have to be taken care of explicitly, and when developing P2P algorithms one needs to play special attention to such limitations.

Another result, that is valid for both the centralised and distributed case, is the observation that certain features of the environment, and in particular, the available data to learn from, are extremely important for selecting the best algorithm. This is a (surprisingly) rather under-emphasised aspect in related work; yet it has a great practical relevance, especially in our applications. One concrete example is the sparsity of the database. In some databases there is an enormous amount of items, such as books, for example, and in other databases there are only a

few hundred. In one database, most of the items are rated by the users, while in other only a very small fraction.

The research has so far focused on the collaborative filtering approach and has not considered the internal features of items, such as natural language (words), category descriptions, metadata, etc. In a pure collaborative setting a sparse database is very problematic, because it is difficult to generalize from patterns of similarity between users. On the other hand, most practical databases can be expected to be very sparse indeed, due to the enormous number of items involved (scientific papers, books, files, etc). For example, the BookCrossing database has these properties. In such cases, a naïve approach to collaborative filtering can provide catastrophic results, for example, worse results than trivial baselines such as predicting every rating with the mean of all ratings. It is therefore very important to use more sophisticated methods such as clustering users, machine learning approaches such as support vector machines, and using more sophisticated probabilistic models such as singular value decomposition and Latent Dirichlet Allocation. Distributed versions of such methods are currently being investigated.

Deviations from the plan and corrective action

The work in WP2.3, which was scheduled to start at month 6, was started by USZ slightly earlier. It was decided that preparatory work, such as surveying the state of the art, could be carried out in WP2.3 from the beginning of the project. The implications of this for the USZ budget and person month allocation are explained in section 6.3.2. Apart from this minor deviation, the objectives have been reached and no impact on other WPs is expected.

Publications

Tölgyesi, N. & Jelasity, M. (2009). Adaptive peer sampling with newscast. In Henk Sips, Dick Epema, and Hai-Xiang Lin, editors, Euro-Par 2009, volume 5704 of *Lecture Notes in Computer Science*, pages 523–534. Springer-Verlag, 2009.

Jelasity, M. (2009). QLectives: Socially Intelligent Systems for Quality. Seminar of the Institute of Informatics, Univ. Szeged. April 2009.

Jelasity, M. (2009). Gossip-based Strategies in Global Optimization. invited plenary talk, Annual meeting of the Hungarian Operations Research Society, Szeged, November 2009.

Tölgyesi, N. & Jelasity, M. (2009). Adaptive peer sampling with newscast. Presented at EuroPar 2009, August 2009.

Stream 3: Empirical datasets collection, processing and validation

The task of Stream 3 is to gather hypotheses (informed by work from Stream 1 and Stream 2) and data-sets from the QLectives platform implementations of the living labs (Stream 4) in order to test models and validate algorithms. First, by contrasting theoretical models with the real systems' behaviour, characterising the level and limits of prediction possible with the given models and suggesting empirically based reformulations. Second, by providing guidelines for implementation derived from the study of models. Stream's 3 contribution to the understanding of social mechanisms will lead to robust quality-emergence in P2P and Web2.0 systems.

First year of Stream 3 was devoted to analyses of the data gathered in so called "Batch 0", i.e. datasets adapted from other projects. The data was translated to standard format and a first analysis has been conducted.

WP3.1 Data collection and experimentation

Contributing partners

UniS, TUD, ETH Zurich, UWAR, CNRS

Objectives for WP3.1

- collecting massive empirical datasets obtained from the two living labs, peer-to-peer systems, social media websites and/or academic paper archives, social rating and bookmarking sites;
- creating an open "living archive" of empirical data for researchers to use; and
- collecting data from specially designed web experiments.

Progress towards objectives

A series of datasets were explored in terms of gathering and assessing the data for batch 0 of the living archive.

- SuprNova.org – swarm level dataset. Including: hash, seeders, leechers (sometimes total downloads, amount of transferred data), download behaviour for 18 months with popularity of files. Includes data on system failures on trackers, web servers, torrent servers, and load balancers,
- Piratebay.org – Covers the 2,000 biggest PirateBay.org BitTorrent swarms which were tracked for a week. Content availability and popularity.
- Filelist.org – Measurement of a private BitTorrent tracker. Data includes: online/offline time connectibility (Firewall/NAT issue), upload/download, upload/download rates, connection time, BitTorrent client.,
- Youtube.com – Public user profile information, published clips, social network information and favourite clips.,
- Del.icio.us – Tagging behaviour, taste diversity, and user activity.
- Flickr.com – dataset1: tagging behaviour, taste diversity (user's favourites), and user activity; dataset2: Flickr Groups dynamics: daily history of variations in population, pool size, moderation features,
- CiteUlike.org – Tagging behaviour, taste diversity, ratings, and user activity,
- LibraryThing.com – Tagging behaviour, friends and interest of book readers,

- Wikis – dataset1: Daily variation in population, number of pages, number of edits (and other variables available on the Special: Statistics page) for the 10K largest Mediawikis (including the largest Wikipedia pages); dataset2: Daily variations of wikis in content, size and structure, represented as a set of global metrics

Most of the databases were identified as not suitable for the research within QLectives, because they contain either too high-level or snapshot-type data. The Wikis and Flickr databases have been used in WP3.3, also the Filelist.org database was chosen for future use as the one having most potential for social network analyses. A Data Archive was established to store all data used in the project in such a way that it would be accessible from the QLectives website.

Where appropriate, it was agreed that the data should be prepared in the following way. The data should be inserted into a normalised MySQL database, as one or more tables, with suitable field types and labels, dumped to a file using MySQLDump, and compressed using gzip. The data should be accompanied by documentation which, minimally, includes date and method of collection, sample (e.g. start and end times and dates, any selection carried out), format of the data (e.g. using the show tables command in MySQL), size of the data file (in bytes, and in number of records per table) and links to articles using the data and other related material.

Significant results

The initial version of the Data Archive containing the "batch 0" databases has been created and can be found at http://www.QLectives.eu/index.php/Batch_0

Other data sets have been found and used by members of QLectives (see WP3.2)

- PGP contact graphs – List of edges of the giant component of the network of users of the Pretty-Good-Privacy algorithm for secure information interchange (data described in: M. Boguña, R. Pastor-Satorras, A. Diaz-Guilera and A. Arenas, *Physical Review E*, vol. 70, 056122 (2004)).
- Email contact graphs – List of edges of the network of e-mail interchanges between members of the Univeristy Rovira i Virgili (Tarragona). (data described in: R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt and A. Arenas, *Physical Review E* , vol. 68, 065103(R), (2003)).
- Epistemic hypergraphs – Bibliographical database on the community studying zebrafish during 1990-2004 and describing hypergraphs where certain nodes are scientists while other nodes are concepts found in abstracts, and hyperlinks are collaboration teams featuring both authors and concepts (data described in C. Roth and JPh. Cointet, *Social Networks*, 32(1):16-29, 2010, and C. Roth SDEAS 1(3), art 2, 2006)

Finally, the partners managing the two platforms that will become QLectives's living-labs (Tribler and Econophysics Forum) have implemented mechanisms to collect data about their users' behaviour. This can be especially interesting from the research application viewpoint in the case of Econophysics Forum, since the platform has been completely redesigned and we expect to be able to observe the whole process of adaptation of 'old' users to the new environment.

WP3.2 Data processing and knowledge extraction

Contributing partners

UniS, TUD, ETH Zurich, UWAR, CNRS, IRT

Objectives for WP3.2

- Summarise and extract measures and statistics from the datasets gathered in WP3.1 in a way that will be relevant and convenient for the work of WP3.3
- Characterise the datasets in order to discover previously unexpected patters, using data-mining and other techniques
- Produce a detailed listing of failures or inefficiencies on existing peer production websites

Progress towards objectives

The data collected and deployed in the QLectives project was agreed to be transformed into a unified format, which was developed in the early stages of the WP3.2.

Significant results

The format was reported in the article Zhang,B., Iosop, A., Pouwelse, J. & Garbacki P. 'A Unified Format for Traces of Peer-to-Peer Systems'. The paper was accepted for the Workshop on Large-Scale System and Application Performance (LSAP), in conjunction with the International Symposium on High Performance Distributed Computing (HPDC 2009), 10th June, 2009, Munich, Germany.

Additionally, the Filelist.org dataset has been converted to the unified format and is available from QLectives server (link to sample of the data: http://www.QLectives.eu/images/9/91/Tom-data.zip)

WP3.3 Model and algorithm validation and prediction (CNRS)

Contributing partners

UniS, ETH Zurich, CNRS

Objectives for WP3.3

- Use data collected and processed in WP 3.1 and WP3.2 to test hypotheses drawn from the theories developed in WP1.1 and the model and algorithms developed in WP2.2 and WP2.3.
- Use data in conjunction with validated models to make predictions about the structure and dynamics of quality collectives

Progress towards objectives

Three of the above datasets have been used for hypothesis validation: wikis, FlickR and epistemic hypergraphs. The first two cases correspond to large-scale studies of the evolution of thousands of wiki websites and FlickR photo-sharing groups. The aim was to check the determinants of website/group growth, both in terms of content and user base. The notion of growth was used as a first proxy for the notion of quality and was correlated with various macroscopic indicators, including population size profiles, structural features (for instance related to the underlying social network) and governance policies. The last case corresponds to an exploration of social and socio-semantic patterns of constitution of academic collaboration teams, where teams are authors gathering to produce a given article at a given time using a given set of concepts and addressing a certain set of topics.

Significant results

The following paper has been accepted for publication in *Scientometrics*: “Academic team formation as evolving hypergraph” by Carla Taramasco, Jean-Philippe Cointet and Camille Roth. This research demonstrates how teams form depending on n-ary attributes, estimating structural and non-structural determinants of group formation rather than dyad formations. It is based on a coevolving socio-semantic framework, through empirical measurements on academic team data: the study describes team formation patterns using an underlying team-based structures and random hypergraphs as null models of collective formation. Another preprint by Lionel Tabourier, Camille Roth and Jean-Philippe Cointet (see WP 1.1) explores the formation of random hypergraphs and shows that several patterns typical of traditional social networks may actually be explained by simple underlying hypergraphic processes.

These *group-level* studies provide original results on teams rather than individuals and are likely to provide robust foundations for subsequent models.

Stream 4: QLectives platform and living lab implementation

The aim of this stream is to produce a deployed Web and Peer-to-Peer infrastructure platform with generic functions for supporting QLectives communities. More precisely, it aims to develop components that are deployable over the infrastructure to provide the specific functions required by the two application domains of QScience and QMedia. In each case mechanisms are applied that promote and facilitate the growth of quality communities (putting the right users together) and the quality of disseminated informational items within the communities (passing the right information to the right people). QScience is currently based on a deployed Web2.0 infrastructure and QMedia is based on a deployed P2P infrastructure. In the following, we give a detailed description of the achievements for each work package separately.

WP4.1 Peer-to-peer platform - P2P-Qual

Contributing partners

TUD, USZ, IRT

Objectives for WP4.1

- Produce a re-usable Peer-to-Peer (P2P) platform infrastructure which facilitates peer-production;
- Craft a zero-server architecture with unbounded scalability, fault tolerance, and high availability;
- Provide generic functions and services required by the QScience and QMedia living lab;
- Implement the world's first zero-server social network system;
- Implement complexity-science inspired techno-social networking fabric primitives:
  - QFlow: augment P2P transfers of information with audit trial of quality indicators,
  - BWCurrency: Transform P2P bandwidth into a transferable currency,
  - P2PWidgets: run-time deployable code units which extend the platform;
- Provide algorithm deployment for Stream 2 and implement collection functions for Stream 3.

Progress towards objectives

This Work Package delivered version 1.0 of the QLectives platform, with a initial sets of re-usable modules that are the base for the implementation of QMedia in WP 4.3 ("Media distribution living lab – QMedia") and may form the core of a future decentralized version of QScience in WP4.2 ("Scientific innovation living lab – QScience").

The QLectives Platform is built on top of the already deployed and mature P2P tribler.org code-base, which provides most of the low-level P2P functionalities for the social networking and quality facilitation required. The software architecture of QLectives Platform is modular so that its components can evolve separately. The QLectives Platform version 1.0 provides the following functionalities:

- For community features and facilitation of quality, it is critical that peers can communicate among themselves, without being prevented by NATs, firewalls, etc. Thus, a NAT-puncturing module has been developed which allows increased connectivity in a P2P system, circumventing prevalent issues in topology management;
- The first version of the P2P-Widgets module prototype allows code to be dynamically loaded from the network in a P2P application, enabling the runtime addition of functionalities to the system.

Significant results

The following section introduces the two most important contributions implemented in the QLectives Platform.

NAT puncturing

NAT puncturing is a technique to increase connectability among peers in a distributed application. In Peer-to-Peer (P2P) systems, computers on the Internet connect with each other in a symmetrical fashion. This requires that arbitrary computers on the Internet are able to connect with each other. However, the deployment of firewalls and Network Address Translator (NAT) boxes creates obstacles. NAT puncturing circumvents some of these obstacles.

By their very nature, firewalls are meant to regulate what connections are permitted. Moreover, firewalls are frequently configured to allow only connections initiated by the computers inside their domain, based on the assumption of a client-server model of communication. Of course, in a P2P setting, connections may also be incoming, often on non-standard ports, which these firewalls don't allow.

NAT boxes pose a separate but related problem. Although NAT by itself is not meant as a connection filtering technology, it does present an obstacle for setting up connections: the publicly visible communications endpoint (IP and port combination) is not visible to the computer behind the NAT box, and may even be different for each remote endpoint. To make matters worse, NAT technology is often combined with firewalling.

The techniques for dealing with NATs and firewalls are well known. For example, the STUN protocol details how a computer can determine what kind of NAT and firewall is between itself and the public Internet. Connection setup can be done through connection brokering or rendez-vous. It should be noted however that the connection setup techniques are most useful for UDP traffic. Setting up a connection for TCP traffic when both computers are behind a NAT/firewall requires non-standard use of TCP and IP mechanisms, and may rely on specific NAT/firewall behaviour to work.

A UDP NAT/firewall puncturing module has been implemented as part of the QLectives Platform. Although operational, the module is not yet used in user downloads at this stage. Because the evaluation of such a tool can only be made when deployed in users' machines, we opted to have the current module test its efficacy and report back its results before we move users' connection processes to rely entirely on this module.

Currently, the NAT puncturing process works as follows. First, the puncturing module builds a swarm, much like the BitTorrent software. However, instead of having a separate tracker, it uses a peer at a pre-programmed address and port and employs a form of Peer EXchange (PEX) to find new peers to connect with. The PEX messages include the peer ID, the IP address and port at which the peer originating the PEX message communicates with the remote peer, and the NAT/firewall type of the remote peer. The latter is included so the receiving peer can determine whether it is useful to attempt to connect to the remote peer, as certain combinations of mapping and filtering are not able to connect to each other.

P2P-Widgets prototype system

P2P Widgets are chunks of code that can be disseminated and dynamically deployed in a P2P network to create new services and applications on top of a P2P substrate. With an efficient method for widget discovery and deployment, users can choose to add new functionality to their

P2P client at runtime, adding for example new group interaction mechanisms or quality assessment code based on their social interactions.

Our requirements for the development of the P2P-Widgets prototype system were the following:

1. Discovery and download of widgets must be done in a scalable, decentralised manner.
2. The widget repository must be browsable, i.e., provide a sorted list of widgets. The list does not have to be complete all the time, but should eventually contain most of the widgets (indeterministic browsing).
3. The widget repository must support rating and reviewing of widgets in a scalable, decentralised manner.
4. The widget system must support local storage per widget.
5. The widget system must support scalable intra-widget communication. The widget system must provide a useful Communication API for intra-widget communication. The widget system must find and maintain a set of intra-widget communication partners.

The following contributions were made:

1. Designing and implementing a fully operational, security enabled P2P-Widget Runtime Environment.
2. Creating a zero-server implementation for discovering and downloading P2P Widgets within a P2P network.
3. Designing and implementing an operational zero-server market system for finding, rating and reviewing P2P Widgets.
4. Extending the widget paradigm with local storage and zero-server intra-widget communication that collaboratively supports global storage.

Publications

Rahman, R.; Hales, D.; Meulpolder, M.: Heinink, V.; Pouwelse, J. & Sips, H. (2009). Robust vote sampling in a P2P media distribution system. Proceedings IPDPS 2009 (HotP2P 2009), IEEE Computer Society.

D4.1.1 Deployed QLectives platform v1

WP4.2 Scientific innovation living lab – QScience

Contributing partners

UniS, TUD, ETH Zurich, USZ, UniFr, IRT

Objectives for WP4.2

- Create a platform to support scientific quality and creativity through proactively linking scientists with shared interests and quality assessments.
- Support rapid formation of new scientific communities with rapid consensus on quality and significance of publications and other scientific resources.
- Grow social capital through reliable reputation and trust mechanisms and thus improve scientific productivity - share tasks, reduce re-invention.
- Collect data from the system for analysis in Stream 3.

Progress towards objectives

In 1998, Fribourg University launched the web-based publication “Econophysics Forum” (EF, see www.unifr.ch/econophysics) which has served as a community ideas exchange platform for the last ten years. Today, it is a leading web site frequently visited and gathering online publications for interdisciplinary and complex sciences, counting over 2,000 registered users. Because it was only

maintained, not upgraded, the web site gradually became outdated according to current IT standards. This concerned not only the page technology but also the page spirit: the opportunity for user contributions was practically non-existent.

The first planned step was to make the EF more open to the input from its users and employ some simple recommendation and community-building tools, while still adhering to the standard Web 2.0 approach (see the manifesto "What is Web 2.0" by T. O'Reilly, http://www.oreilly.com/go/web2, for detailed discussion of this approach). A considerably upgraded version of the EF has been prepared and available to users since autumn 2009. It is still in a testing regime but its new functions will be advertised to users in spring 2010.

Further work has focussed on preparing a roadmap towards a vision based on an open source platform containing core functionality that could be instantiated by several scientific communities in parallel, but that would maintain interoperability at several levels such as user profiles and metadata about the articles, as well as user ratings, so that a group-evolutionary dynamics could be supported allowing quality products and quality groups to emerge.

Significant results

The following section introduces the most important new functions implemented at the Econophysics Forum (see the screenshot below for an illustration).

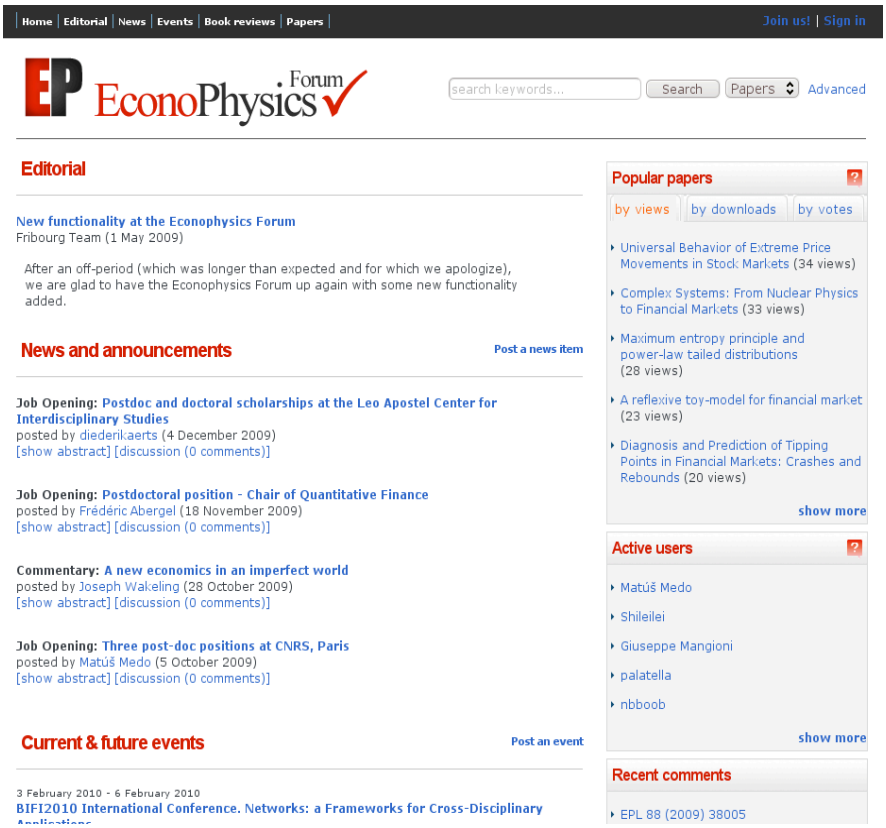
User accounts: Users can now register and create accounts. With a user name and password, a user can log in and use all the features available on the web page. Even users without an account (so-called guest users) can see all the content on the page and can contribute new content. The only limitation is that posts by unregistered users, before being displayed on the page, are placed in a queue awaiting editorial approval, which can cause a few days' delay in their visibility to other users. In this way, the Econophysics Forum is both secure from uncontrolled content (input by spammers or otherwise malicious users) and at the same time open to input from those who are not willing to spend time creating a new account.

Users' input: Users can now contribute directly to each section except for editorials (where guest editorials are still welcome but, before being posted, are approved by the editors). In a few clicks, any registered user of the EF can post a news item (job opening, link to a relevant popular article or blog, and others), information about a coming event (workshop, conference, etc.), book review, or a scientific paper. Besides submitting new content, users can also express their opinion about the existing content by comments and votes. Finally, to judge the quality of discussion comments (which can be interesting pieces of knowledge themselves but also can be inappropriate, unfair, or even malicious), users can give plus or minus rating to each individual comment.

Statistics and trends: The new version of the EF traces the behaviour of its users and generates some simple statistics and trends from these data. The results are shown directly on the front page in three distinct information boxes on the right hand side (see figure below) in a similar way as it is nowadays done on many information portals (compare with, for example, www.cnn.com).
Personal tools: All user names are links pointing to the personal pages of these users. This personal page includes the user's e-mail address, link to a homepage, affiliation, and history of the user's contributions (submissions, comments, and votes). Users can opt not to show this information by appropriate settings in their personal profiles. In contrast to the very basic information available about other users, each user's own personal page includes a personalized recommendation that is not visible to the others.

This recommendation includes a list of papers recommended to the given user and a list of users similar to the given user (in this way, one can look for potential collaborators or simply for people to discuss some work with).

Figure 1. New front page of the Econophysics Forum.



Deviations from the plan and corrective action

Data collection from the QScience living lab (the Econophysics Forum) is delayed as a result of: (a) prolonged testing phase of the web page, and (b) problems with the informatics service of University of Fribourg (reluctance to provide complete web logs in particular). As the testing phase is now completed, specific needs will be communicated to the informatics service in as short time as possible. The data collection will then be initiated, accompanied with a public announcement of newly implemented features of the Econophysics Forum.

Publications

D4.2.1 Deployed QScience living lab v1 (Web2.0)

WP4.3 Media distribution living lab - QMedia

Contributing partners

TUD, USZ, IRT

Objectives for WP4.3

- Create an experimental next-generation user centric and social media distribution platform (QMedia);
- Build upon the P2P-Qual platform and offer seamless download and streaming of open media content;
- Provide social networking functionality allowing users to freely share media with friends and contacts;
- Self-organise, through detection of taste similarity and user rating systems, groups of similar peers;
- Implement social tools allowing users to create a shared viewing experience;
- Promote the discovery of high quality open content by integrating the WP4.4 quality meta-data system;
- This work package deals with the development of the QLectives Platform.

Progress towards objectives

This Work Package delivered version 1.0 of the QMedia platform, addressing an initial set of the required functionalities for this living lab, namely a new and user-friendly graphical user interface, a micropublishing service, a user interface for the interaction with the P2PWidgets infrastructure developed in WP 4.1 (“Peer-to-peer platform - P2P-Qual”) and proofs of concept widgets that use the infrastructure and user interface.

As the QLectives Platform, QMedia also builds on the Tribler code base. Tribler is an open source media-sharing client, which has been extended to meet the QLectives requirements. The implementation towards version 1.0 of QMedia concentrated on the following developments:

- CommentCast: a decentralized micropublishing service built around media-sharing channels;
- New user interface: the existing user interface in Tribler was considered a major obstacle for growing a large user base for QMedia, and was therefore redesigned from scratch to provide a user-friendly and intuitive experience;
- User interface for the P2PWidgets and examples of P2PWidgets: P2PWidgets allow the dynamic deployment of functionality over the peer-to-peer network formed in QMedia; in this work package we have implemented a user interface to expose the functionality

represented by any widget to the user and to allow users to publish and discover widgets. To demonstrate this interface, we also implemented some widgets.

Significant results

The following section introduces three main modules developed for QMedia v1.0.

CommentCast

The huge popularity of the Twitter micropublishing service has recently caught the attention of the media and of the research community. In a general sense, micropublishing can be seen as a process where each source publishes certain content and is followed by a (usually different) number of followers who may also respond to the published content. Both the source and the published content can be diverse: it could be an everyday user updating his daily life to his friend(s) or a commercial company broadcasting a new release of its product to its clients. Both friends and clients can respond to the published content by posting some comments. In this way, a timeline of the source is established.

CommentCast is a first design of a decentralized micropublishing service for QMedia. This service gives each participant in the QMedia network the possibility to post and reply to comments towards a Channel, which is an abstraction for a media source in QMedia.

Each participant in the system keeps a private and a subjective shared comment history, which saves comments posted by this participant and by his peers. Periodically, this peer exchanges a subset of these two comment histories with other peers. The period and candidate peers are decided by BuddyCast, the underlying gossip protocol of QLectives Platform. Considering security issues, peers only spread the comments belonging to the channels to which they have subscribed (peers only subscribe to a channel when they believe that channel is in good quality). In this way, the more popular (in terms of number of subscriptions) the channels are, the faster their comments are spread, and low quality channels will die off eventually, together with and due to those malicious comments.

New user interface
The success of the QMedia living lab depends on having a large number of loyal users in the deployed system. This observation, in turn, makes the usability of the system paramount.

The elicited requirements for the user interface are for it to be appealing and easily usable by a broad audience with minimal assumptions about their technical knowledge and skills. The approach chosen to make the overall user interface more appealing was to revamp the look-and-feel of the whole graphical user interface, giving it a more vivid and modern mood. As for usability, we opted to replace Tribler's user interface into a new one that was as simple as possible, relying on the widespread known metaphor of a search engine portal, and to optimize for the central use-case in the system: search, acquire and watch.

There is now one central element of interaction that is exposed to the user: the search box. This search box is similar to that exposed by a number of other systems that virtually all users of the Web routinely rely on, such as Google, Yahoo! and Bing. There are two other components in the interface of this window: a sharing bar that indicates how much the user has shared with his peers compared to how much he has downloaded from them, and a set of links in the right upper corner that allow the user to navigate to other tabs of the interface, which represent the different contexts in which the user might act. Furthermore, the interface redesign also led to a change in the colours used in the interface, relying on white to refer even more to the Web and search design contexts already familiar to users, and light colours to evoke a vibrant and friendly atmosphere.

Besides the searching context, there are three other contexts in which the user can interact with QMedia. The Settings tab allows a user to configure the software in a user-friendly manner, and with the My Files tab the user can revisit files downloaded in the past and follow the progress of downloads he has started.

Thirdly, a Channels tab is implemented; a channel is a series of files with quality asserted by a user, and represents a first step towards the implementation of a wider social experience in QMedia. Channels are meant to help users discover quality content amidst the sea of available files; to collectively accredit users capable of identifying such quality as hubs of community activity and to eventually create digital spaces of shared interest that can seed group interaction and lead to community formation.

In the present version of QMedia, each user has his own channel. Channels are discovered by two means. The first option is for the user to find them similarly to finding files: users select the type of search they are performing in the drop menu next to the search box and use the same mechanism as when searching for files. The second option for channel discovery is to explore channels based on popularity.

User interface and examples for the P2PWidgets

The infrastructure for dissemination, acquisition and deployment of P2PWidgets was developed in Work Package 4.1. In this work package, the user interface of QMedia was extended to include P2PWidgets and examples implemented that serve as proofs-of-concept.

The interface to the widgets is built around the Widget Market. This market is a widget itself that allows users to browse widgets based on popularity, to search for specific widgets, install discovered widgets, and to publish, rate and review widgets.

To demonstrate the system's capabilities, several widgets have been created. They are described in detail in Deliverable D4.3.1.

Publications

Meulpolder, M., D'Acounto, Capotă, M., Wojciechowski, M., Pouwelse, J.A., Epema, D.H.J., Sips, H. J. (2010) Public and private BitTorrent communities: A measurement study. Accepted for IPTPS 2010, San Jose, California, USA.

Hales, D. (2009). Tribler. Invited talk at the European Conference on Complex Systems 2009, COSI-ICT'09, University of Warwick, Coventry, UK, 21-25 September 2009.

WP4.4 Quality search and discovery

Contributing partners

TUD, USZ, IRT

Objectives for WP4.4

- Bring public service broadcasting quality commitment to the YouTube and BitTorrent generation of media distribution
- Explore ways for sharing quality meta-data from, and with, public broadcasters
- Give users the ability to rate, moderate, annotate and recommend contents themselves
- Allow for easy "injection" of existing open meta-data sources (i.e. inter-operate with open standards)
- Increase quality by reducing spam, incorrect or malicious meta-data via application of novel mechanisms from Stream 1 and 2.
- Hence create a global distributed self-organising high quality Electronic Programme Guide (EPG)

Progress towards objectives

The work in WP 4.4 started with an analysis of existing and standardised metadata formats, such as TV-Anytime (TVA) and MPEG7, since those are the ones currently adopted by public broadcasters. QMedia needs to interface – amongst other potential metadata formats coming from the Web2.0 and the Peer-to-Peer domains - to the metadata-frameworks currently in use in the broadcast domains.

After having carried out a survey on and an assessment of existing metadata models to be used for representing the QLectives-specific representations of Quality, it was concluded that existing and accessible schemes do not meet the requirements of the project. They were either not generic, or not specific and flexible enough. Handling, processing and collection of metadata in QLectives may result in a complex scheme, and an adequate generic metadata architecture is considered to be helpful, especially when it comes to designing algorithms to calculate and express Quality by means of metadata.

Requirements coming from existing rating, profile building and recommendation systems were then analysed. The QLectives-specific recommendation system should not solely be based on "just" a rating system, but also incorporate all kinds of existing metadata. For example, interpretation of available technical parameters like video resolution, bitrate, and type of deployed codec might give information about visual quality. Other technical metadata aspects are viewing statistics, e.g. how often content is consumed completely or partially. Another question is how to design a rating system that is attractive to users to give their votes, in particular, what exactly and to what extent user should be compelled to rate. In addition, a high quality recommendation system will incorporate ratings about the raters themselves to reduce the risk of introducing spam ratings. Consequently, the model needs to be capable of expressing rating and individual evaluation of data about any subject or object, such as the creator of a digital item, as well as the algorithm that performed the rating calculation.

With respect to system design and implementation, we consider a generic metadata scheme that can be adapted to existing metadata sources and formats to be the only acceptable approach for achieving this. A generic metadata model in UML (Unified Modelling Language) was developed, representing a platform- and implementation-independent model/scheme in terms of software

design or software architecture, expressed as class schemes. It should form the basis of developing appropriate algorithms and enables QLectives to use existing metadata types, like the media-optimised TV-Anytime standard, as well as possibly entirely new, QLectives-specific or other types of metadata yet to be defined, without the need for defining another complete scheme. The work is reported in deliverable D4.4.1

Additionally, work has been launched on filtering algorithms, being part of Deliverable D4.4.2 “Algorithms for detecting and handling HQ-MD (High Quality Meta Data) and Spam metadata”. A first application of the generic metadata framework has been performed, involving generating a specialisation of the so-called “MetadataQualifier” class, and developing class models and use cases. The iterative development of QLectives-specific applications should be regarded also as an interface between the metadata/content and algorithms. The task has also a strong link to WP2.3 “Algorithms for trusted and quality rankings”, and a workshop has been organised in March 2010 between the relevant partners to take the research further.

Significant results

With the generic metadata model, we have a toolkit available with respect to further work to be done in WP4.4, for example, allowing an easy “injection” of existing open meta-data sources.

Publications

D4.4.1 Generic metadata model

4 Deliverable and milestone tables

The following table lists the deliverables due to be delivered during the first twelve months of the project.

Del. No.	Deliverable name	Lead beneficiary	Nature	Dissemination level	Delivery date from Annex I	Delivered (yes/no)	Actual/ delivery date (month)
D1.1.1	Overview of theories and models of complex techno-social systems	ETH Zurich	R	PU	9	Yes	9
D1.3.1	Simulation of quality emergence	UF	R	PU	12	Yes	12
D2.1.1	Candidate theory models for cooperation algorithms	TUD	R	PU	12	Yes	12
D4.1.1	Deployed QLectives platform v1	TUD	P	PU	12	Yes	12
D4.2.1	Deployed Qscience living lab v1 (Web2.0)	UF	P	PU	12	Yes	12
D4.3.1	Deployed Qmedia living lab v1	TUD	P	PU	12	Yes	12
D4.4.1	Generic metadata model	IRT	P	PU	12	Yes	12
D5.1	Functional project website	UniS	O	PU	12	Yes	12
D5.3	Report on dissemination activities	UniS	R	PU	12	Yes	12
D6.1	Annual report	UniS	R	PU	12	Yes	12

Table 1: Deliverables

The following table lists the milestones due to be achieved during the first twelve months of the project.

Milestone	Milestone name	WP number	Lead beneficiary	Delivery date from Annex 1	Achieved (Yes/No)	Actual achievement date
M1	Completion of data collection (batch 0)	WP3.1, WP3.2	ETH Zurich	6	Yes	31.8.2009
M2	Completion of data processing and analysis (batch 0)	WP3.2, WP3.3	UWAR	9	Yes	30.11.2009
M3	Theory and algorithm reformulation (batch 0)	All WP's in Stream 1 and 2	UniS	12	Yes	28.2.2010

Table 2: Milestones