# ModelMMORPG - D2.3.- Report on data analysis

**Schatten, Markus; Okreša Đurić, Bogdan**

**Other document types / Ostale vrste dokumenata**

*Publication year / Godina izdavanja:* **2015**

*Permanent link / Trajna poveznica:* https://urn.nsk.hr/urn:nbn:hr:211:689967

*Rights / Prava:* Attribution-NonCommercial 3.0 Unported/Imenovanje-Nekomercijalno 3.0

*Download date / Datum preuzimanja:* **2025-03-07**

SVEUČILIŠTE U ZAGREBU
FAKULTET ORGANIZACIJE I INFORMATIKE
VARAŽDIN

*Repository / Repozitorij:*

Faculty of Organization and Informatics - Digital Repository

DIGITALNI AKADEMSKI ARHIVI I REPOZITORIJI


# Large-Scale Multi-Agent Modeling of Massively Multi-Player On-line Role-Playing Games

## D2.3.- Report on data analysis

This project was funded by the Croatian Science Foundation

Principal investigator:

**Markus Schatten**

lab

2015.

Copyright © 2015 Artificial Intelligence Laboratory

PUBLISHED BY ARTIFICIAL INTELLIGENCE LABORATORY,
FACULTY OF ORGANIZATION AND INFORMATICS, UNIVERSITY OF ZAGREB

HTTP://AI.FOI.HR/MODELMMORPG

Licensed under the Creative Commons Attribution-NonCommercial 3.0 Unported License (the "License"). You may not use this file except in compliance with the License. You may obtain a copy of the License at http://creativecommons.org/licenses/by-nc/3.0. Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

Some of the results presented in this deliverable have been previously published in [20].

*Technical Report No. AIL2015003 – First release, September 2015*

*Document compiled by: Bogdan Okreša Đurić, Markus Schatten with inputs from other project team members*

*This work has been supported in full by the Croatian Science Foundation under the project number 8537.*

# Contents

# 1. Project Description

## 1.1 Abstract

Massively multi-player on-line role playing games (MMORPGs) give us the opportunity to study two important aspects of computing: (1) large-scale virtual social interaction of people (players) and (2) the design, development and coordination of large-scale distributed artificial inteligence (AI). A common denominator for both aspects are the methods used to study them: social interaction can be descibed and simulated using agent-based models (ABM social science perspective) whilst distributed AI is commonly modelled in terms of multi-agent systems (MAS computer science perspective).

The important question to ask in both perspectives is how do agents organize in order to perform their tasks and reach their objectives? Project ModelMMORPG (Large-Scale Multi-Agent Modelling of Massively On-Line Role-Playing Games) will employ a combined empirical and theoretical approach towards finding the answer to this question.

From the empirical side, we shall study the human behaviour on a number of venues across various gaming servers in order to find most suitable structures, cultures, processes, strategies and dynamics employed by most successful player communities. From the theoretical side, we shall test a multitude of organizational architectures from organization theory in various MMORPG settings, and compare them with methods found in empirical research.

Our research is therefore aimed towards enriching the organizational design methods for the development of MMORPG to foster the development of self-organizing and adaptable networks of large-scale multi-agent systems.

With this in mind, our main goals are:

1. To identify and formalize adequate organizational design methods for developing LSMAS in MMORPGs.
2. To couple them with real-life and future scenarios from industry.
3. To provide open and accessible tools, which will allow for design, development, implementation, control, simulation and maintenance of LSMAS in MMORPG

## 1.2 Introduction

Role-playing video or computer games (commonly referred to as only role-playing games or RPGs) are a game genre in which the player controls the actions of some protagonist (or potentially several party members) in a world which is well defined [27]. A massively multi-player on-line game (MMOG) is a (computer) game that supports a great number of players playing on-line simultaneously causing or even fostering interaction among them [25]. Massively multi-player on-line role playing games are thus a mixture of these two genres allowing players to control the action of their protagonist (avatar) by interacting with a potentially large user-base on-line [26].

The global market for MMO games is growing rapidly with 2011 $\approx$ 8.5 billion €, 2012 $\approx$ 10.2 Bn€, 2013 $\approx$ 11.7 Bn€ and 2014 $\approx$ 15.0 Bn€ [6, 12]. While the economic importance of MMORPGs is obvious, another aspect is of equal importance: it allows us to investigate two aspects of large-scale computing - (1) social interaction of (large numbers of) players through a computing platform as well as (2) the design and implementation large-scale distributed artificial intelligence (in form of non-player characters – NPCs, mobs – various monsters to be fought, as well as AI players – bots). Both aspects can and should be studied using agent-based methods, the former by ABM (a social science perspective) and the latter by MAS (a computer science perspective), whilst the important question to ask in both perspectives is: how do agents organize in order to perform their tasks and reach their objectives?

The ModelMMORPG (Large-Scale Multi-Agent Modeling of Massively On-Line Role-Playing Games) project will employ a combined empirical and theoretical approach towards finding the answer to this question. From the empirical side, we shall study the human behavior on a number of venues across various gaming servers in order to find most suitable organizational structures, cultures, processes, strategies and dynamics employed by most successful player communities. From the theoretical side, we shall test a multitude of organizational architectures from organization theory in various MMORPG settings, and compare them with methods found in empirical research. The research is therefore aimed towards enriching the organizational design methods for the development of MMORPGs and to understand the undelying principles of self-organizing and adaptable networks of large-scale multi-agent systems.

MMORPGs have a number of different subgenres, but a usual setting is that a protagonist is placed into a world in which he interacts with various NPCs and mobs which give out tasks (quests) that it has to solve to be able to buy better equipment, learn new skills like magic and similar, or proceed to higher levels. In ModelMMORPG we have chosen The Mana World (TMW)[1] MMORPG to conduct our research. The reasons for selection were: (a) it is open source (GPL licensed) allowing us to modify code and add additional functionality, (b) it has a supportive community, (c) it supports a number of interaction techniques which can be studied (e.g. trade among players, IRC based chat, organizing teams called parties, social network functions e.g. friends, enemies, parties etc.), (d) it is a (more or less) finished game featuring lots of quests that can be analyzed.

In order to answer our outlined questions, we firstly designed a special quest in which players ought to organize their activities in order to solve it. The quest is designed during a 3-day brainstorming session. Later on, during a data collection phase we will allow players to play the game in order to collect their behavioral data during a period of one month. Additionally to real players, we will add AI players that shall play alongside humans in order to provide a baseline for a future (organized) version of AI players. After data collection, the data will be analyzed using social network analysis (SNA) and natural language processing (NLP) techniques in order to find patterns of organizational behavior among successful players. After these patterns have been formalized, we will try to use them for building organized agents (AI players) and repeat

[1] See `http://themanaworld.org` for details.

the experiment in order to see if such agents behave better or worse then the initial non-organized version. In this deliverable we will particularly report on the quest design and implementation as well as the (initial) AI player implementation.

## 1.3 Team Members

**Markus Schatten** (Principal investigator)
Head of Artificial Intelligence Laboratory
markus.schatten@foi.hr
+385 42 390891

**Joaquim Belo Lopes Filipe**
Departament of Systems and Informatics, Escola Superior de Tecnologia de Setúbal
joaquim.filipe@estsetubal.ips.pt
+351 265 790 040

**Nikola Ivković**
Department of Computing and Technology
nikola.ivkovic@foi.hr
+385 42 390872

**Mladen Konecki**
Department of Theoretical and Applied Foundations of Information Sciences
mladen.konecki@foi.hr
+385 42 390873

**Mario Konecki**
Department of Theoretical and Applied Foundations of Information Sciences
mario.konecki@foi.hr
+385 42 390834

**Robert Kudelić**
Department of Theoretical and Applied Foundations of Information Sciences
robert.kudelic@foi.hr
+385 42 390852

**Ivan Magdalenić**
Department of Computing and Technology
ivan.magdalenic@foi.hr
+385 42 390872

**Marko Maliković**
Faculty of Humanities and Social Sciences Rijeka
marko@ffri.hr
+385-(0)51-265-765

**Bogdan Okreša Đurić**
Artificial Intelligence Laboratory
dokresa@foi.hr
+385 42 390891

**Jurica Ševa**
Department of Theoretical and Applied Foundations of Information Sciences
jseva@foi.hr
+385 42 390873

**Pietro Terna**
Università di Torino
pietro.terna@unito.it
+39 011 6706067

**Igor Tomičić**
Department of Computing and Technology
igor.tomicic@foi.hr
+385 42 390869

# 2. Introduction

The social web has become a major repository of social and behavioural data that is of exceptional interest to the social science and humanities research community.[1] Computer science has only recently developed various technologies and techniques that allow for harvesting, organizing and analyzing such data and provide knowledge and insights into the structure and behaviour of people on-line. Some of these techniques include social web mining, conceptual and social network analysis and modeling, tag clouds, topic maps, folksonomies, complex network visualizations, modeling of processes on networks, agent based models of social network emergence, natural language processing, opinion mining and sentiment analysis. In the following chapter we will brifly introduce all of these techniques, give example studies and ideas on applying some of these techniques on the study of Massively On-Line Role-Playing Games (MMORPG). MMORPGs include a number of characteristics which make it social application: (1) large numbers of people interact on them, (2) they feature various communication facilities including chat, instant messaging, user forums etc., (3) they feature social networking through the establishment of guilds, parties etc. or through direct establishment of ties between players (friends, enemies, blocklists etc.). Thus we will approach MMORPGs as social applications when trying to analyze them.

The world wide web and especially the social web represents a vast repository where 'millions of people, on thousands of sites, across hundreds of diverse cultures, leave "trails" about various aspects of their lives' [16, p. 13]. If we, for a moment, adopt the interpretation of Niklas Luhmann that social systems are systems of communication and only communication [8] we might find a justification that by analyzing the "trails"1 we can analyze the social system itself: politics, culture, media, business, technology, science etc.

As it has already been outlined in our previous studies [11, 24] there are numerous more or less automated methods and techniques dealing with various aspects of information organization and retrieval that can be applied and used in social science research. Due to the fact that the social web has tremendously influenced the social discourse and lots of "trails" about social standpoints can be found on all over the world wide web, the use of advanced big data, web mining as well as semantic web and agent based techniques seems to be a fruitful direction for future studies.

[1] Parts of this chapter have been published in EQPAM [19, 20]

## 2.1 (Social) Web Mining

In order to get to know the discipline of web mining, it is advisable to grasp the basic concept of data mining, a more general approach to data. Data mining, in its broadest term, can be defined as a process of extracting knowledge, especially informative knowledge, from a large collection of data. Web mining is considered a more specialised version of data mining, i.e. "web mining is the means of utilizing data mining methods to induce and extract useful information from web data information" [28]. It is evident that domain of web mining lies heavily in the world wide web area, and that the final product is information, or, better yet, informative knowledge based on and extracted from, the provided data.

Typically, web mining is divided into three types describing different aspects of web data and different techniques they use: web content mining, web structure mining and web usage mining. According to Ting, "web content mining analyzes content on the web, such as text, graphs, graphics, etc" [23] wherefore the main technology used in this type of web mining is natural language processing, described in further detail below. "Web structure mining is a technique to analyze and explain the links and structure of websites" [23]. The most suitable concept for use in this technique, and the most convenient theory, is network theory, also described later. The main interest of this type of web mining is in creation of computer programs, called crawlers and harvesters, with the main goal of automated extraction and construction of websites' structure. Lastly, "web usage mining can be used to analyze how websites have been used, such as determining the navigation behavior of users" [23] . This type of web mining is the one which is connected the most with social and psychological sciences, since it is mainly concerned with users and their behaviour while using services of the Web.

It is self-evident that the web survived great development since its start, let alone in the last couple of years. Nowadays, with the maturity of web 2.0 technologies, which allow users free and uninhibited creation and publication of data on the web, where this same data becomes instantly available to a multitude of users, web is becoming an invaluable source of social data.

Social networking applications are a part of this mass data creation trend, successfully evolving a special instance of web mining, i.e. social web mining. Ting defines social networking as a concept "usually formed and constructed by daily and continuous communication between people" [23], therefore including many different social relationships, e.g. closeness among individuals or groups. Experience, thought and feelings sharing has always been incorporated in the life of humans, hence it is no wonder that the age of web 2.0 uncovered on-line social networking as arguably the most massively used feature of the web. Rapid growth of different types of communication which provide good platforms for users to communicate and share data, incorporating on-line social networking into everyday lives, include photo sharing services (e.g. Flickr), video sharing services (e.g. Vimeo), professional networking services (e.g. LinedIn), full fledged social networking services (e.g. Facebook), blogging services ( e.g. Wordpress), microblogging services (e.g. Twitter), instant messaging services (e.g. Skype) and many others. All these services made interpersonal communication, relationships and human behaviour readily available on-line, creating easy-to-use datasets for analysis. According to Ting, the "history of social networks analysis is [...] dating back more than a hundred years to around the 1900s" [23], and mostly in the field of sociology. One might argue that scientists conducting experiments around the 1900s could not predict the amount of data available today. Computing power used nowadays is capable of analysing and reasoning on not only data provided by services mentioned above, but content in comment sections of news portals and similar services, popular wiki pages and huge collections of data, all available on the Web.

Although data is available, analyzing it is not a job easily done. Several different types of data exist, demanding corresponding method of analysis to be used: structured data, usually formatted in tables and easily read can be analyzed in a manner similar to an ordinary database (e.g. exchange rates); semi-structured data, including text data widely available on the web and containing a lot of information, demands improved methods of analysis including NLP (e.g. blogging services); unstructured data, mostly consists of images, video or audio files, and web applications (e.g. flash applications).

The world wide web allows for adding a temporal dimension to the data, and, by extend, to the analysis. Collecting data through time creates a dataset which contains indications on how a certain trend developed through time. For example, following a certain Twitter hashtag during the time of presidential elections in Croatia in 2014, one could have conducted an elaborate analysis to depict the state of the people in the country and abroad, thus following popularity of one candidate over the other, grading and predicting their actions through time. Adding geographical dimension to the temporal data, it might be possible to follow an event through time and space. For example, analysing the Giro d'Italia bicycling event tweets, it might be possible to follow the race in real time, following comments spectators make. Similarly, following presidential candidates on tour before elections might result in interesting insight in their actions, and the effect those actions have on the people.

Analysing data with all its components (e.g. temporal and geographical) can lead to information which can make easier following and keeping check of your voters in elections. Furthermore, the gathered information can make you more successful in influencing them and knowing what exactly, and where, is a topic of great importance or of more interest to the local population. 2012 U.S. presidential election candidates worked hard on their social networking propaganda, influencing youth, but gaining information on other parts of population as well.

There is a potential problem though, in the context. Data is indeed beneficial if one is sure of the subject the data is about. Should the data lack context, analysis becomes a tough problem. Statements might become sarcastic, or lose their sarcasm. Simple sentences could have their meaning inverted. For example, should one not know the nature of a satirical portal, one might make conclusions based on the content of an article, misinterpret and miss the general idea completely. Therefore, it is very important to know the context of data creation, publication and consumption.

## 2.2 Social & Conceptual Network Analysis

The 'new science of networks' as Barabasi likes to call network theory, allows us to study networks regardless of their origin. In the context of social web analytics, two types of networks are of particular interest: social and conceptual. While the former represent linkages (communication, friendship, interaction, trade, cooperation etc.) between social entities (people, organizations, social groups, countries, political parties etc.), the latter provides insights into the structure (synonymy, mutual context, homonymy, hyperlink etc.) and dynamics (evolution of context) of concepts (words, ideas, phrases, symbols, web pages etc.).

Both social and conceptual networks are ubiquitous on the social web. For example, apart from the obvious that two people are connected if they are friends on a social networking site like Facebook, two people might be connected if they have commented on the same topic on some forum, have liked the same article or video on some news portal or podcasting site, have bookmarked the same web page or are subscribed to the same news feed. As one can see, if this topic, article, video and/or news feed has a political context we might argue that these two people have a similar political interest or attitude. If we now multiply this situation on hundreds or thousands of users, we can use social network analysis techniques (like finding connected components or clustering) to identify subnetworks of people that form groups of similar political mentalities. Depending on the particular criteria of network formation (the context of the observed network) we might form hypotheses about the actual criteria. For example, if the criteria is a video about a political speech of some political actor, we might find that some speeches of different actors actually have the same or a similar group of interested followers. Additionally, if we observe such social networks over time we might also reason about the dynamics an evolution of groups: how and when did they form? Having such data available, would allow us for example, to develop agent based models that might predict this observed behaviour.

Apart from being analyzed using SNA methods, social network can be visualized to provide appealing, yet informative insights into a social system. For example, in a study we have visualized a social network of co-authors on a ICT related conference over the years (see figure 2.1). While such visualizations can be beautiful, they provide us with important information: in this case we can see that a certain core of authors participates with a paper every year with more or less the same co-authors, while the majority of authors participates only once and never comes back. This allows us to draw certain conclusions about this conference and might indicate to the conference organizers that some things should be changed.

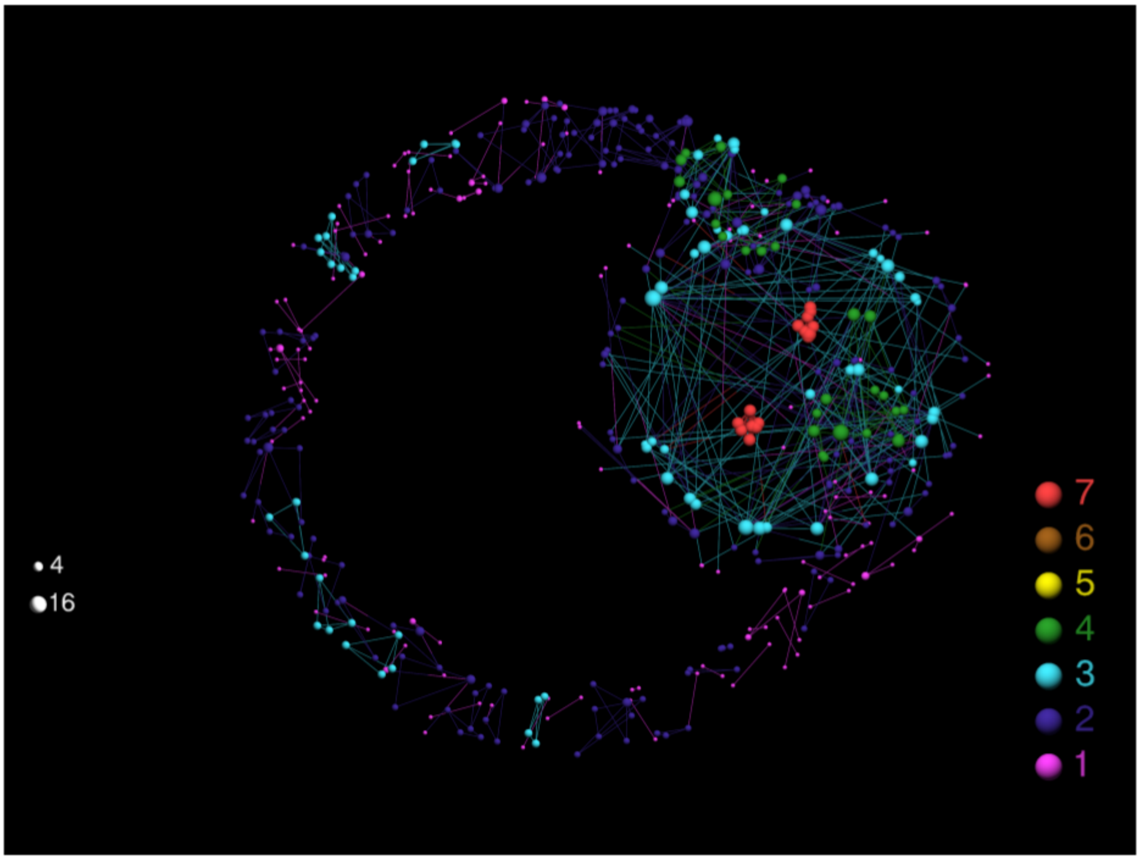


Figure 2.1: Social network of co-authors from an ICT conference [21]

Conceptual networks, on the other hand, aren't maybe that obvious: two web pages where one links to the other, two keywords used on the same article, to tags used on the same on-line video, two extracted1 concepts which appear in the same text etc. Such conceptual networks can be studied in a number of ways: in the following we will show four such techniques (tag clouds, topic maps, folksonomies and visualizations), but there are multitudes of other usable methods.

### 2.2.1 Tag Clouds

Various tagging applications are pervasive across social media. On most blogs users are allowed to tag content with keywords they believe accurately describes the content. These tags can be considered to be certain meta-data about the actual content: users who have read / reviewed the content have in a way preprocessed it and provided us with keywords that can give us important insights about it without even reading it. Especially, when this process is multiplied, e.g. when multitudes of users tag the same content, due to a certain collective intelligence these tags provide an accurate descriptor of the content.

On a lot of systems that use social tagging, so called tag clouds [22] are implemented which are simple visualizations of the tags provided by users. Usually, the most used tags are shown in a bigger font, while less used tags are smaller. Some applications like Wordle1 even have basic NLP capabilities and are able to auto-summarize text in form of beautiful keyword clouds.

While keyword clouds are not in the common sense a conceptual network analysis method, they still allow us to create a visual representation of text and context that provides an insight into the matter with a simple look on it. We have used this technique in a number of studies to visualize important keywords on research papers. For example figure 2.2. shows a keyword cloud which was assembled by analyzing research papers dealing with data mining applications in tourism. From this visualization, one can easily comprehend which were the most important approaches, methods and technologies used in a great number of studies.


Figure 2.2: A keyword cloud visualization of data mining in tourism related papers [2]

Similar keyword clouds can be constructed for arbitrary sets of articles with keywords / tags which might also be generated through NLP techniques. As an example we have implemented a platform which has been used for web mining, keyword cloud visualization as well as conceptual network analysis of the Croatian Scientific Bibliography (CROSBI).1 The implemented system, which is still work in progress, used a number of technologies including web scraping (Scrapy),2 an object-relational database (PostgreSQL)3 and an advanced scripting language (Python).4 The selection of technologies wasn't arbitrary, Scrapy allowed for easy implementation of a number of harvester agents that collected and extracted data from semi-structured documents that were stored in a specially designed PostgreSQL database. PostgreSQL was selected due to its unique text mining and NLP capabilities that allowed for automated dictionary based text stemming. PostgreSQL uses dictionaries to eliminate words that should not be considered in a search (so called stop words), and to normalize words so that different derived forms of the same word will match (lexemes). Python was used to glue this technologies together and provide analysis related features.

CROSBI is a social application in which Croatian scientists provide bibliographic data about their publications (Schatten, 2013). A usual entry includes authors, title, type of publication, abstract, keywords, link to document (if available), language, databases the publication is abstracted in, scientific field, category, as well as a number of additional fields depending of document type like journal name or publisher.

### 2.2.2 Topic Maps

Topic maps are a more advanced conceptual network visualization technique. Topic maps show the development of topics in a certain textual discourse through time. The more often a certain concept was used in some text, the greater surface on the topic map. Peaks show when a certain concept was most often used.

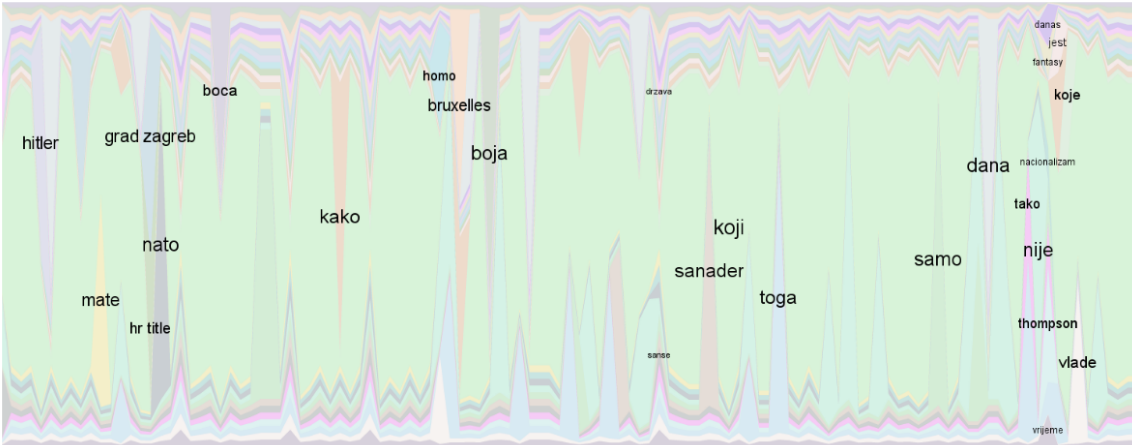


Figure 2.3: A topic map from a major Croatian political blogging site [15]

For example, figure 2.3. shows a topic map generated from data harvested from a major Croatian political blogging site.1 It was used to identify most important concepts in the given time-frame, as well as the time-line of the topics that were analyzed.

While in this context topic maps were used on on-line (social blogging or forum etc.) text, they are applicable to any series of text that have a common spatio-temporal frame. For example, one might take a number of political newspaper articles from a given time frame (e.g. the 80's) from a given space (e.g. Yugoslavia), automatically extract keywords (using NLP techniques for example) and then visualize the development of topics.

### 2.2.3 Folksonomies

The term folksonomy coming from folk and taxonomy was coined by Mika (2007) to label a certain mathematical structure in a special context. He was analyzing the social bookmarking site Delicious on which users were able to bookmark web pages they encountered and add keywords to describe each page. These keywords were then merged from all users and used to facilitate a "socially powered" search engine. What Mika has observed was that each bookmark has three components: (1) a user who made it, (2) a web page that is being tagged, and (3) a keyword that describes the page in the mind of the user.

This observation gave rise to the idea to model the set of data obtained by bookmarking as a tripartite hypergraph (which represents the folksonomy) in which every (hyper)node consists of the three outlined components. By using a procedure called graph folding it is possible to construct bipartite graphs and moreover to connect keywords based on various criteria and thus form conceptual networks. For example, two keywords can be considered to be connected if they have been used on the same web page, and likewise if they have been used by the same user. The former criteria has been shown to yield an intuitively well structured conceptual network of connected concepts.

This conceptual network can further be clustered to yield well connected components as shown on figure 2.4. As one might see, these components show what where the major themes of web pages users of Delicious bookmarked: free time, web design, sex, business and travel.

While this folksonomy model seems to be specific for bookmarking sites, we have shown in [21] and [18] that it is applicable to almost any social content related data like a bibliography for example. Also, the model is not constrained on only three dimensions, but any number of

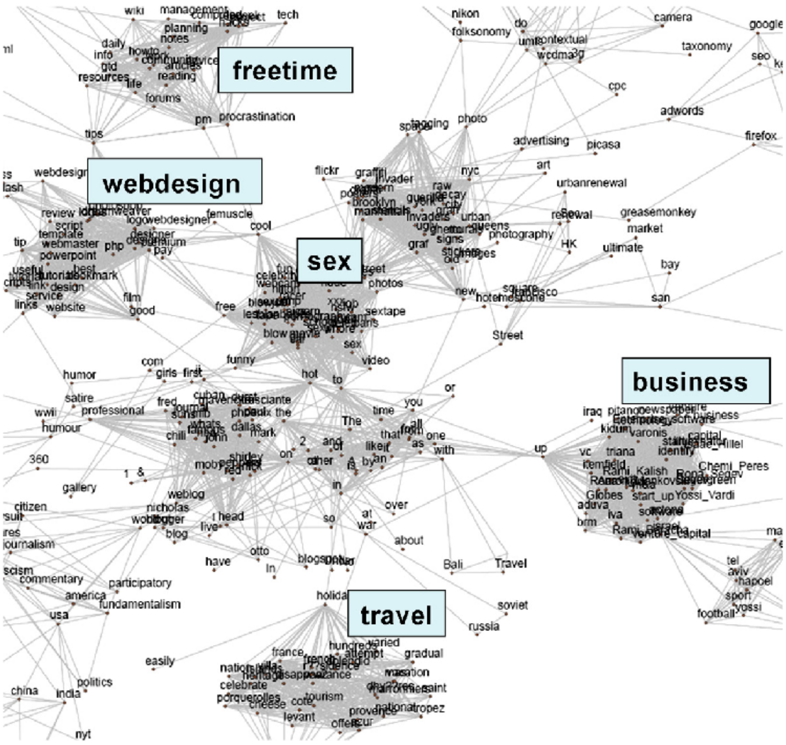


Figure 2.4: Folksonomy visualization of a social tagging system called Delicious [9]

dimensions can be used depending on the actual dataset. For example in [18] we have used 4 dimensions to analyze the Croatian scientific bibliography - each hypernode consists of (1) an author who has participated in writing some scientific article, (2) a keyword provided by the authors, (3) the actual article, and (4) the scientific field the author(s) have categorized the paper.

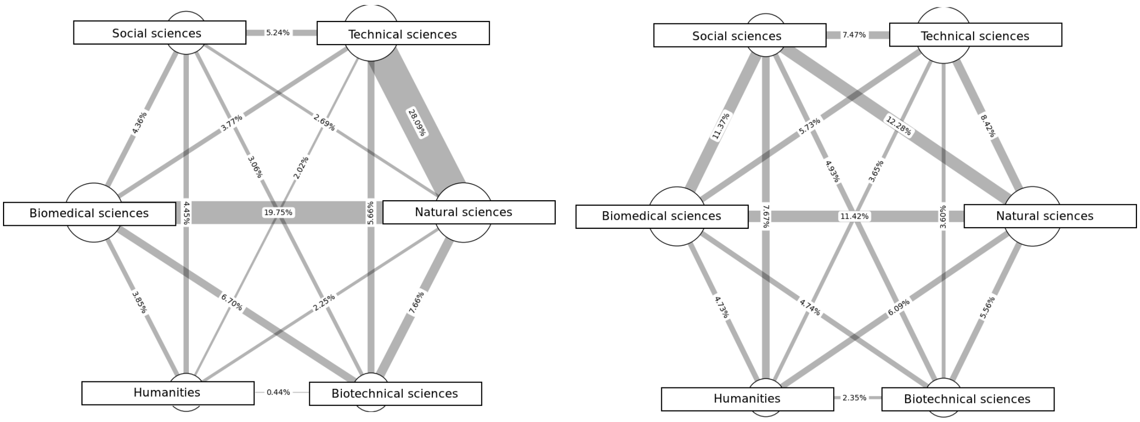


Figure 2.5: Connections among scientific fields in Croatia (left - social criteria, right - conceptual criteria) [18]

The last dimension allowed us to compare the connection between scientific fields on a global level according to two criteria: (1) social (scientific collaboration) and (2) conceptual (same keywords are used in both fields). Figure 2.5. shows both such obtained networks and indicated discrepancies: e.g. some fields are mutually conceptually well connected but there was only little collaboration between them.

As one can see, the folksonomy model can be used to analyze any repository of tagged text: bibliographies, news articles with keywords, tagged blog entries, categorized wiki entries etc. The "tags" can be keywords provided by authors or other people, categories, labels, geographical or other origin, field of study, titles, topics etc.

### 2.2.4 Complex Conceptual Network Visualizations

To come back to the study outlined in the previous subsection, the constructed conceptual networks of the different scientific fields were quite complex, with hundreds of thousands of nodes. Since such data sets are hard to represent analytically or in form of tables, the field of complex network visualization has emerged.

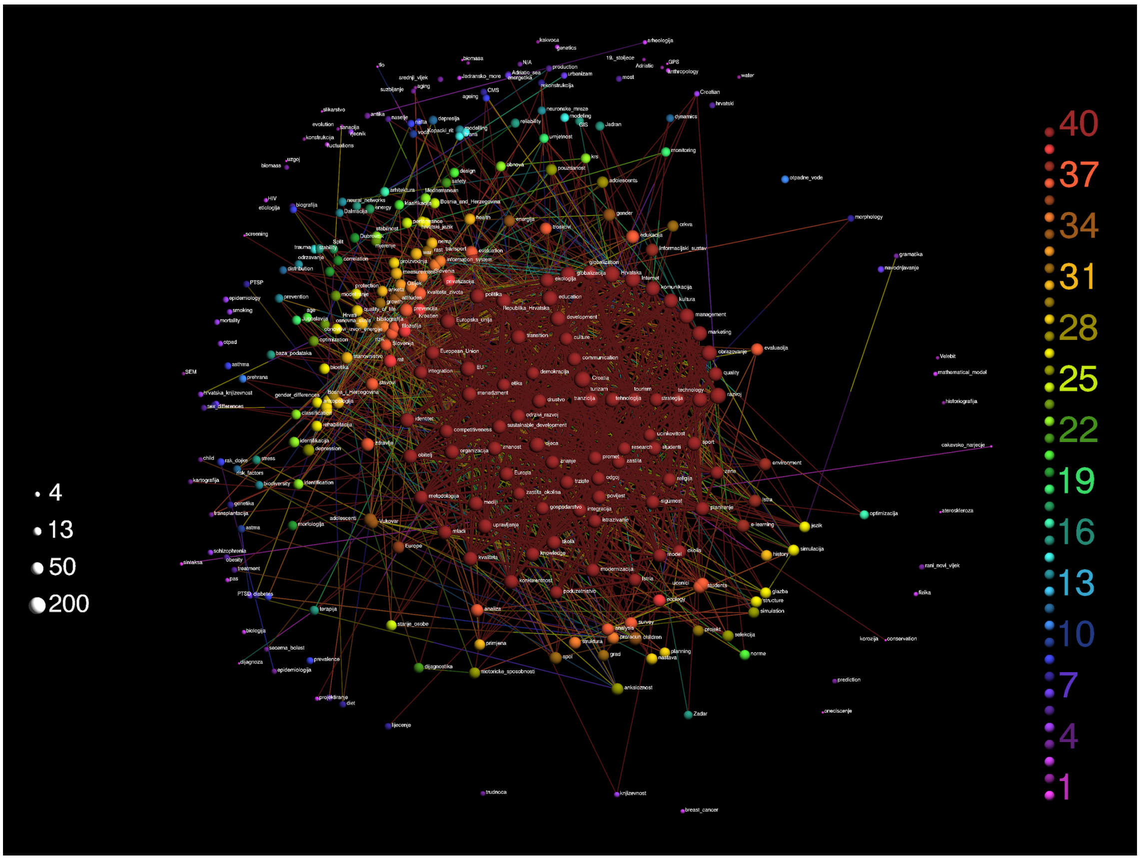


Figure 2.6: Complex conceptual network visualization of most important concepts in the social sciences according to an analysis of the Croatian Scientific Bibliography [18]

The visualization of complex networks is an art and a science. Visualizations have to be appealing, informative and concise. Numerous algorithms for complex network visualization have been proposed and developed, each of which adjusted to different kinds of networks. For example, on figure 2.6. we have used the k-core decomposition algorithm to visualize the conceptual network of keywords used in the social sciences on the Croatian scientific bibliography. What emerged in the red inner core, are the most important concepts the social sciences have dealt with in the target time frame.

### 2.2.5 Processes on Networks

While the previous examples more or less exclusively dealt with the (static) structure, one needs to mention that there are also methods that allow us to study the dynamics and evolution of networks.

One particular type of methods that we want to point out here are virus spreading models [14] which describe the mathematics of how viruses spread through a networks of people. While there seems to be no obvious connection to social science, we need to point out that a very similar mechanism is behind the spreading of rumors [10] and information [29].

By using empirical data acquired through web mining one could develop agent based models that allow us to understand the spreading of certain themes that we want to study. For example, if we want to model a process in which some political attitude towards a certain political actor has spread through a social network, we might first harvest the communication data between people on a particular network (for example a forum, a blogging site or some news portals which allow user commenting). Then we extract only those texts (e.g. messages, articles, blog posts etc.) that deal with the political actor in question as well as temporal data (time of publication). We might then employ advanced analysis techniques to (1) construct the social network of people involved in the discourse, (2) identify the different attitudes towards the actor by using NLP techniques or sentiment analysis, and (3) define the actual process of information spreading in form of an agent based model for example.

### 2.2.6 SNA & ABM

A method different from those mentioned above enables analysis of a distinct set of, primarily social, entities and their interaction, is reachable by using combined power of social network analysis (SNA) and agent-based modelling (ABM), since this coalescence allows "embedding a huge amount of data in user-friendly models" [4]. These models ease information retrieval from the provided data through techniques of both SNA and ABM. The role of ABM is modelling agents and their interaction based on mathematical and area-specific models, whilst the role of SNA lies in unveiling the structure of interaction of the modelled agents and efficiency and stability of the network. The benefit of using SNA and ABM together, according to Fontana & Terna, is that one takes care of the problems of the other, e.g. SNA creates serious problems with exploring possible sets of nodes' configurations, since it can only be achieved by the means of combinatorics, producing an exponential number of possible sets, yet ABM makes this problem evanish, on account of the number of agents in a model being limited only by computational power. Fontana & Terna stress that the number of possible configuration remains enormous, but it is possible to mitigate this problem.

In order to successfully represent the idea of combining SNA and ABM, Fontana and Terna devised recipeWorld - "an agent-based model that simulates the emergence of networks out of a decentralized autonomous interaction" [4]. The idea is that modelled agents are given recipes (variable number of steps to be taken in order to achieve a given end), and "they are activated, following their internal rules and capabilities, by the events, and the network emerges as a side effect" [4].

This proposed model is based on four distinct sets: (A) is the actual world populated by entities and their actual network; (B) is an ABM with agents which base their behaviour on the orders and recipes derived from (A); (C) represents the network generated by (B), as represented in Figure 2.7.

Fontana and Terna propose the possibility of populating sets (C), (B) and (A), respectively, by knowing (D), representing known data on the network, and using inference and a sort of reverse engineering, as shown in Figure 2.8.

Combination of the above approach with the aforementioned methods, such as web mining and network analysis, presents one with the possibility of using this reverse engineering approach and creating agent-based models based on gathered social web data. For example, social web mining might provide one with enough data (D) to model a political network (C) showing e.g. a number of political parties and their interaction, or some other social entities, including voters, political figures, numerous governmental and non-governmental organizations, societies, or even countries.

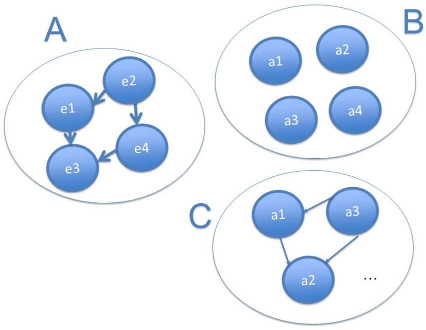


Figure 2.7: Sets A, B and C [4]

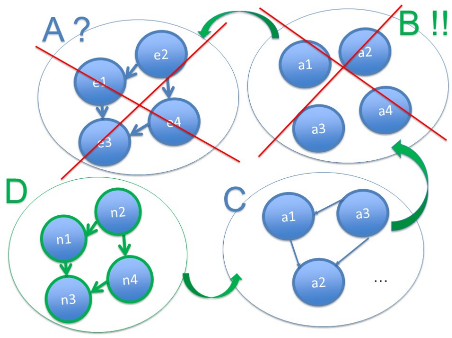


Figure 2.8: Process of obtaining A, B, and C by knowing D [4]

In turn, this network (C) can serve as the basis for an ABM, (B), hence allowing for simulation of real-world entities and their actual network (A).

## 2.3 Speech Recognition & Computer Vision

Since the social web does not only include textual data, but also multitudes of audio, video and image data, in order to analyze it in an automated or semi-automated fashion, one needs to employ advanced computing techniques.

Speech recognition is a technique that transforms recorded audio of human speech to text. This can be used to transcript various audio tapes and videos that include human speech, identify different speakers and then use the above mentioned techniques to analyze it. While speech recognition is a well developed technology, there is a serious drawback of using it to study specific materials: it is in most cases language dependent, which means that for each language one wants to analyze, one needs and adequate speech to text system for that particular language. If there is no such system available, and there often isn't for some languages, one needs to develop such a system which is a long and difficult undertaking.

Computer vision is a field of computer science that aims on allowing computers to analyze image and/or video data in order to recognize objects that are on the given image/video. As opposed to the previous one, these methods aren't language dependent, but are even worse object dependent, which means that for every type of object one wants to recognize, there needs to be an adequate system which recognizes it. This is especially true for person recognition or biometrics (see [1]) where one needs to train the system to recognize every single person of interest (this process is called enrollment). Still, there are some types of objects which can more or less be recognized by even simple system like text on an image or in a video. Such text can then be used in conjunction with the recognized speech.

## 2.4 Natural Language Processing

Natural Language Processing (NLP) can be defined as a process of making a computer system understand the meaning behind written and/or audio data that contains language based information. To give a more comprehensive definition one can quote [3] which states that NLP "is normally used to describe the function of software or hardware components in a computer system which can analyze or synthesize spoken or written language". The goal of systems that analyze and process textual or audio data in a given natural language is to achieve a level of language understanding that is characteristic for humans and their language processing abilities. The main problem in achieving this goal is the complexity of the phenomenon that is the target of NLP: natural language. One of the examples is the word 'bank': it can be a financial institution, a river shore, relying on something etc. The other problem is that language itself is a living entity that evolves during time. The language itself consists of grammar ( the logical rules of combining words of the language to make comprehensive sentences that have meaning) and the lexicon (the words and the situations of their usage). There are specific linguistic tools that make it easier for the algorithms to access, decompose and make sense of a sentence already available for use in processing NLP data. Those tools are as follows [3]:

- Sentence delimiters and tokenizers – detecting sentence boundaries and determining parts of sentences (based on punctuation characters)
- Stemmers and taggers – morphological analysis that links the word with a root form and word labeling giving information if a word is a noun, verb, adjective etc.
- Noun phrase and name recognizers – labeling the words with the noun phrase (e.g. adjective + noun) and recognizing names
- Parsers and grammars – recognizing well-formed phrase and sentence structures in a sentence

The entire process can be broken down in to several steps. In general, the entire text is first broken down into paragraphs, paragraphs into sentences and sentences into individual words that are then tagged (to recognize parts of speech among other) before the parsing begins. The full suite of tools available are sentence delimiters, tokenizers, stemmers and parts of speech taggers, but they are not used in full in all situations. The role of sentence delimiters and tokenizers is the determination of the scope of the sentences and identifying the members of a sentence. Sentence delimiters try to find the boundaries of a sentence, which can be a hard task since the usual sentence endings (e.g. period) can represent other meaning. They are usually created by using expression rules. Tokenizers segment a list of characters into meaningful units that are called tokens. Creating tokens is language dependent as there are differences in how different languages mark word breaks (Latin based languages use white space as a word delimiter). They are usually created using rules, finite state machines, statistical models and lexicons. Stemmers are used to find out the root form of a word. There are two types of stemmers:

- inflectional, that express the syntactic relations between words of the same part of speech (focus is on grammatical features such as present/past or singular/plural)
- derivational, that try to bind different words to the same root (e.g. kind/unkind share the same root). They are supported by the use of rules and lexicons (they relate any form of a word to its root form).

Parts of speech (POS) taggers label the grammatical type of a word, assigning labels and deciding if a word is a noun, a verb, an adjective, etc. Since the same sentence can have more than one meaning, often POS taggers tag the same word with more than one label. POS taggers can be

- rule based, that rely on linguistic knowledge to rule out tags that are syntactically incorrect, or
- stochastic, that rely on training data and assign tags based on frequency probabilities (they are computed from the supplied training set that was matched by hand).

Although POS taggers are very useful in determining the structure and type of words in a sentence, some tasks require their more specific use. At that point, noun phrase parsers and name recognizers are used. Their goal is to identify major constituents of a sentence (e.g. a name).

The history of NLP and it's development has seen several main approaches in implementing previously mentioned tools. The early 1950's, which are regarded as the decade where NLP started,

- Symbolic approach
- Statistical approach

Symbolic Approach is based on the existence of a predefined grammar that is then used in data analysis. The grammars consist of known grammatical patterns and their meaning which are then looked upon in the analyzed data. Due to the limitations of grammar based approach, researchers started to look at the possibilities of other approaches. The statistical approach is based on machine learning and data mining and this approach tries to use available data and allow for the algorithms to learn based on them. The basis of this approach is the existence of a corpora which is, in fact, a hand annotated with the appropriate linguistic labels (e.g. noun, first name, person, etc.) or concepts (e.g. person, building, organization etc.) and an learning algorithm which then uses the corpora and tries to "learn" new concepts from an unlabelled data set based on the rules extracted from the corpora. This approach itself has also seen a change in it's approach. Primarily, the approach was based on decision trees (as similar to the grammars used in symbolic approach)which are derived from a number of if else statements in order to analyze the data to modern, statistical approaches where each recognized concept is defined through a weighting scheme/system. This approach is the basis for all modern NLP algorithms, although their combination is also a frequent approach.

As previously stated, the main current approach to NLP is the use of statistical through machine learning and data mining algorithms. Although this approach has many advantages over symbolic approach one has to have a corpus, a previously annotated data set used as the basis for the algorithms used in the implemented algorithms' learning phase. As it is to be expected, the languages with large linguistic support are languages that have many users (e.g. english). Central and eastern European languages are, unfortunately, still largely unsupported as the necessary tools for a good computational language understanding are still either missing or highly undeveloped. The overall support for central and eastern-european languages is given in Figure 2.9 and 2.10. A good starting point for these language is the META-NET repository [7].

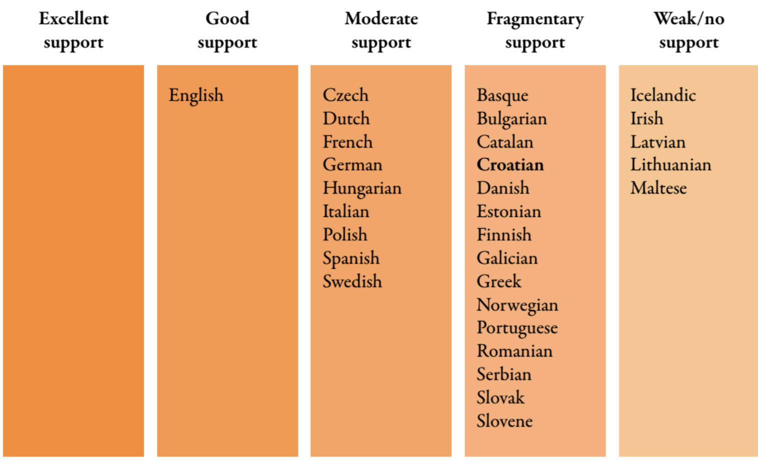


Figure 2.9: Speech and text resources: state of support for 30 European languages [7]

NLP can find its application in numerous analyses inluding but not limited to finding most important concepts in some corpora of text, auto-summarization, automated classification of documents and similar.

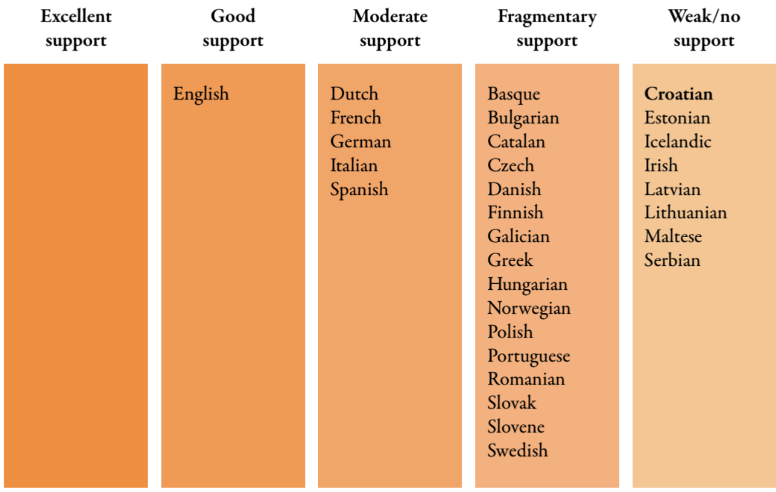


Figure 2.10: Text analysis: state of language technology support for 30 European languages [7]

## 2.5 Opinion Mining & Sentiment Analysis

Sentiment analysis or opinion mining refers to the application of NLP, computational linguistics, and text analytics to identify and extract subjective information in source materials (see [13]). This amazing technology allows us to identify the sentiment (e.g. the feelings) a person had while writing some text or message.

There are numerous application areas of sentiment analysis including but not limited to: (1) product/services research for marketing purposes, (2) better search engines, (3) campaign monitoring (e.g. political, business, showbiz, etc.), (4) recommendation systems (see below), (5) detection of "flames", (6) public health (e.g. emotional/mental status, detecting potential suicide victims etc.), (7) government intelligence, (8) trend monitoring (e.g. social, cultural, political etc.) and similar.

The possible use of sentiment analysis in the study of political attitudes seems to be obvious: for example one might want to study the sentiments a social group has had towards a given political actor over some period of time. What one needs to do is: (1) harvest all communication data from one or more social media applications or other sources like newspapers and similar in the given time frame, (2) filter the data and find only those that contain comments on the actual actor that we want to analyze, and (3) feed this data to the sentiment analysis tool (or tools) for the given language to obtain sentiments over time.

Again, similarly to speech recognition and NLP, sentiment analysis is language dependent. In order to use sentiment analysis on a given text, there has to be a sentiment analysis database or system for the given language. Sadly, there are no sentiment analysis systems for soe languages, for example only recently there were initiatives to implement a Croatian sentiment database [5, 17]. What also needs to be mentioned here is that types of sentiments in various systems are limited: often only positive or negative (with various degrees between these to polarities) are recognized. Complex sentiments are hard to detect, and thus often not implemented or an active field of study. Additionally, since opinion mining applications have a potentially big commercial value, they are often not freely available.

# 3. Relations

In the following chapter we will present an analysis of the data collected during a 2 month period on the developed ModelMMORPG test-bed. Herein we have chosen to analyze the social networks that have emerged between the players while playing the quest for the dragon egg. The data analysis is mostly performed using Pajek[1].

## 3.1 Character Relations

Every character in The Mana World can be related to another in-game character using one of the following relations actions:

- Be Friend,
- Disregard,
- Ignore,
- Blacklist,
- Set as Enemy,
- Erase.

Data analysis performed after the testing period was over generated a network of characters which had developed various relations. 96% of these character relations were friendly, leading us to conclude that users used character relations as a mechanism for keeping in contact with their in-game friends and monitoring their on-line presence. Generated network visualization is show in Fig. 3.1.

Network visualized in Fig. 3.1 shows only characters who were engaged in some kind of relation with another in-game character. Four weak components are clearly recognizable, and are shown using differently coloured nodes. Size of a node depends on the in-degree of the node. Some characters have no other characters which would initiate a relation with them, therefore they have in-degree equal to zero, and are presented by no circle, only a label. In-degree is chosen to be represented here because of its meaning in forming relations - a character can feel something for another character, but this feeling does not have to be returned. Therefore, it is interesting to observe how often characters received affection and friendly relation, even if they did not anything in return. Moreover, as mentioned earlier, in-degree of a node can in this case be observed as a number of other characters who find it useful to monitor whereabouts of the related character,

[1] http://mrvar.fdv.uni-lj.si/pajek/

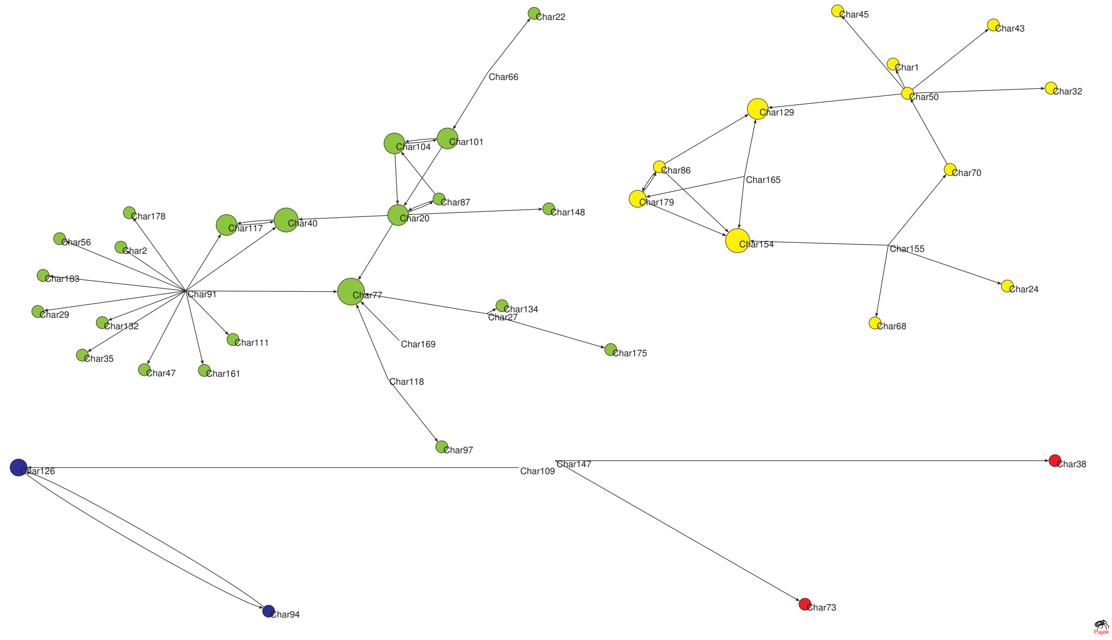


Figure 3.1: Network visualization of character relations with coloured components and nodes sized according to their in-degrees

e.g. when they are online and where (in- game) they are positioned. Nodes are labeled using their in-degree ($k_i^{in}$) and a random number as follows: $[k_i^{in}]\ N_i$ where $N_i$ is a random number such that $1 \leq N_i \leq 47$.

### Strong components

Four strong components consisting of at least two nodes are identifiable. These strong components of two or more nodes are shown in Fig. 3.2 below. Node size here depends on the sum of in- and out-degrees of each node.

### Closeness Centrality

Average all closeness centrality in the network, according to Pajek, is $0,1562$. Closeness centrality per node is shown in Tbl. 3.1 below.

### Degree Centrality

Average all closeness centrality in the network, according to Pajek, is $0,1562$. Closeness centrality per node is shown in Tbl. 3.2 below.

## 3.2 Character Parties

In order to succeed in the game, players must engage in cooperation with other players. Most obvious result of such behaviour is creation of parties, i.e. groups of player characters. Such groups allow for easier cooperation and teamwork, e.g. group chat, experience sharing, etc.

Several parties were formed during the testing period. Since the game mechanics designed for this project demanded cooperation realized using parties of at least three members, only such parties are of interest. All parties are shown in Fig. 3.3, with parties containing $\leq 2$ members coloured light blue, and parties of interest (nodes numbered 49, 48, 75, 14, 43, and 62) coloured

| Node | In | Out | All |
|---|---|---|---|
| Char68 | 0,0426 | 0,0000 | 0,0993 |
| Char47 | 0,0426 | 0,0000 | 0,1965 |
| Char45 | 0,0426 | 0,0000 | 0,1076 |
| Char43 | 0,0426 | 0,0000 | 0,1076 |
| Char66 | 0,0000 | 0,0870 | 0,1450 |
| Char117 | 0,0701 | 0,0426 | 0,2197 |
| Char111 | 0,0426 | 0,0000 | 0,1965 |
| Char178 | 0,0426 | 0,0000 | 0,1965 |
| Char86 | 0,0426 | 0,0851 | 0,1291 |
| Char87 | 0,0532 | 0,0917 | 0,1778 |
| Char175 | 0,0426 | 0,0000 | 0,1556 |
| Char155 | 0,0000 | 0,1114 | 0,1434 |
| Char154 | 0,1064 | 0,0000 | 0,1249 |
| Char132 | 0,0426 | 0,0000 | 0,1965 |
| Char179 | 0,0638 | 0,0638 | 0,1106 |
| Char134 | 0,0426 | 0,0000 | 0,1556 |
| Char77 | 0,1368 | 0,0000 | 0,2987 |
| Char38 | 0,0426 | 0,0000 | 0,0426 |
| Char73 | 0,0426 | 0,0000 | 0,0426 |
| Char70 | 0,0426 | 0,0812 | 0,1434 |
| Char32 | 0,0426 | 0,0000 | 0,1076 |
| Char35 | 0,0426 | 0,0000 | 0,1965 |
| Char165 | 0,0000 | 0,0851 | 0,1291 |
| Char183 | 0,0426 | 0,0000 | 0,1965 |
| Char56 | 0,0426 | 0,0000 | 0,1965 |
| Char50 | 0,0426 | 0,1277 | 0,1613 |
| Char101 | 0,0608 | 0,0917 | 0,1867 |
| Char40 | 0,0993 | 0,0426 | 0,2449 |
| Char104 | 0,0709 | 0,0917 | 0,1844 |
| Char91 | 0,0000 | 0,2979 | 0,2929 |
| Char118 | 0,0000 | 0,0638 | 0,2046 |
| Char109 | 0,0000 | 0,0426 | 0,0426 |
| Char94 | 0,0426 | 0,0426 | 0,0426 |
| Char97 | 0,0426 | 0,0000 | 0,1524 |
| Char129 | 0,0894 | 0,0000 | 0,1383 |
| Char161 | 0,0426 | 0,0000 | 0,1965 |
| Char148 | 0,0638 | 0,0000 | 0,1757 |
| Char1 | 0,0426 | 0,0000 | 0,1076 |
| Char147 | 0,0000 | 0,0638 | 0,0638 |
| Char2 | 0,0426 | 0,0000 | 0,1965 |
| Char169 | 0,0000 | 0,0426 | 0,1991 |
| Char126 | 0,0638 | 0,0426 | 0,0638 |
| Char29 | 0,0426 | 0,0000 | 0,1965 |
| Char24 | 0,0426 | 0,0000 | 0,0993 |
| Char27 | 0,0000 | 0,0851 | 0,2104 |
| Char20 | 0,0851 | 0,1083 | 0,2489 |
| Char22 | 0,0426 | 0,0000 | 0,1167 |

Table 3.1: Closeness centrality of the character relations network by node and closeness centrality type

| # | In | Out | All |
|---|---|---|---|
| 1. | 2 | 2 | 4 |
| 2. | 5 | 0 | 5 |
| 3. | 1 | 1 | 2 |
| 4. | 1 | 5 | 6 |
| 5. | 1 | 0 | 1 |
| 6. | 1 | 0 | 1 |
| 7. | 1 | 0 | 1 |
| 8. | 0 | 3 | 3 |
| 9. | 2 | 2 | 4 |
| 10. | 1 | 0 | 1 |
| 11. | 1 | 0 | 1 |
| 12. | 1 | 0 | 1 |
| 13. | 3 | 4 | 7 |
| 14. | 4 | 0 | 4 |
| 15. | 1 | 0 | 1 |
| 16. | 1 | 0 | 1 |
| 17. | 3 | 1 | 4 |
| 18. | 0 | 2 | 2 |
| 19. | 2 | 1 | 3 |
| 20. | 1 | 1 | 2 |
| 21. | 1 | 0 | 1 |
| 22. | 1 | 0 | 1 |
| 23. | 1 | 0 | 1 |
| 24. | 1 | 2 | 3 |
| 25. | 1 | 0 | 1 |
| 26. | 0 | 13 | 13 |
| 27. | 0 | 1 | 1 |
| 28. | 1 | 0 | 1 |
| 29. | 1 | 0 | 1 |
| 30. | 1 | 0 | 1 |
| 31. | 1 | 0 | 1 |
| 32. | 1 | 0 | 1 |
| 33. | 2 | 2 | 4 |
| 34. | 0 | 4 | 4 |
| 35. | 1 | 0 | 1 |
| 36. | 3 | 0 | 3 |
| 37. | 1 | 0 | 1 |
| 38. | 1 | 0 | 1 |
| 39. | 0 | 2 | 2 |
| 40. | 1 | 0 | 1 |
| 41. | 0 | 3 | 3 |
| 42. | 0 | 1 | 1 |
| 43. | 1 | 3 | 4 |
| 44. | 1 | 0 | 1 |
| 45. | 0 | 2 | 2 |
| 46. | 2 | 1 | 3 |
| 47. | 1 | 0 | 1 |

Table 3.2: Degrees of the character relations network by node and degree type

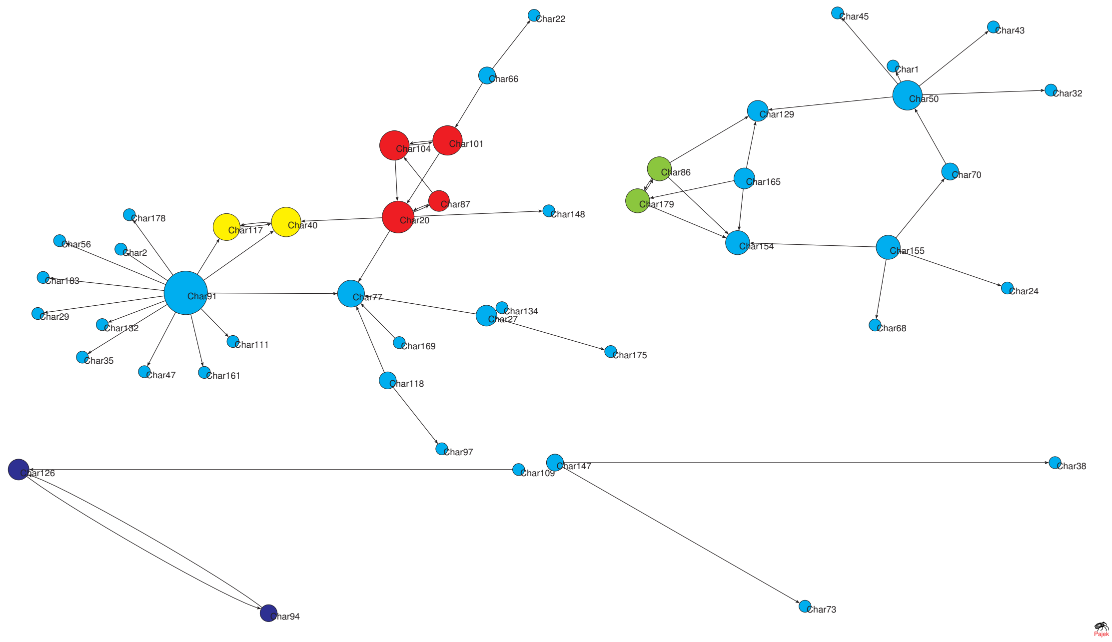


Figure 3.2: Network visualization of character relations with coloured strong components

distinctively. Node sizes are related to out-degrees of particular nodes, therefore party-node size is defined by number of party members.

## 3.3 Party Members and Social Relations

Adding character social relations to the network of party members allows for analysis of social relations between party and non-party members. Naturally, some grouping occurs. Arcs representing social relations have $1 \leq weight \leq 6$ with weight of 1 for friendly relation, and arcs representing party membership have weight 7. Pajek detected 14 islands by line weight with three to ten members each, as visible in Fig. 3.4. Party membership is shown using purple arrows and social relations are shown using red arrows.

Communities formed by Louvian clustering method using default parameter values are depicted in Fig. 3.5.

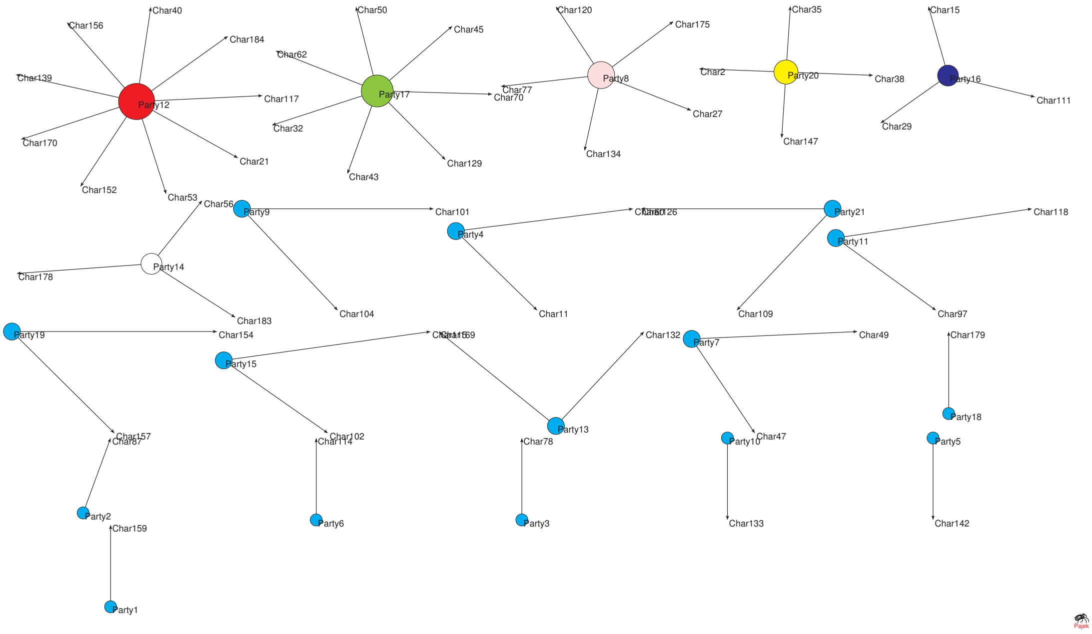


Figure 3.3: Character parties with their members; parties of interest (nodes numbered 49, 48, 75, 14, 43, and 62) are coloured distinctively: white, dark blue, pink, yellow, green, and red, respectively

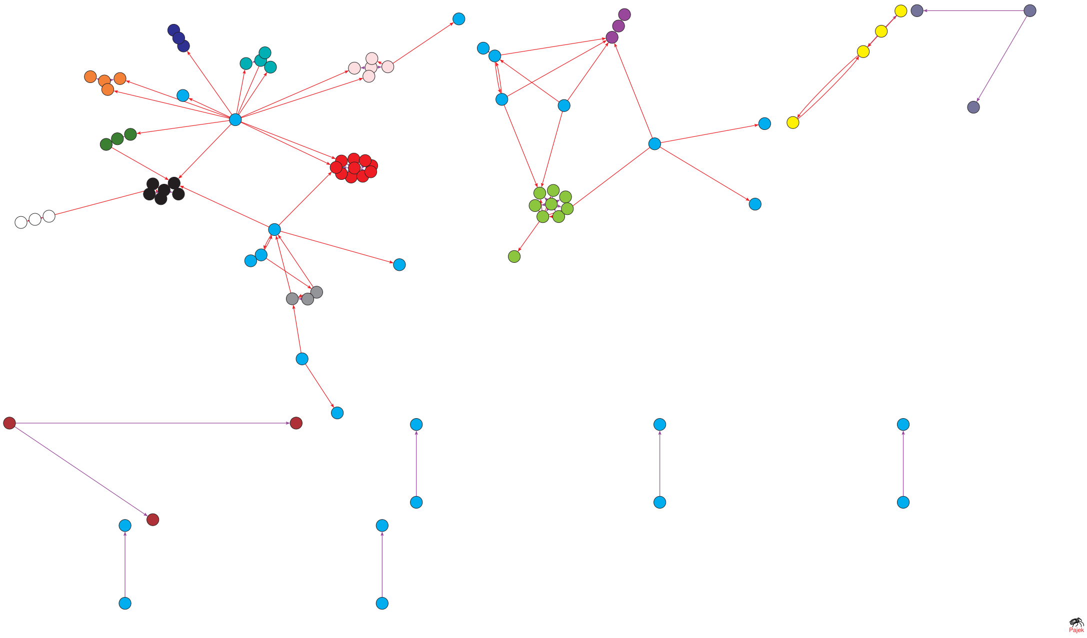

Figure 3.4: Character social relations (red) and party membership (purple) formed islands of characters, which are grouped mostly according to their party allegiance

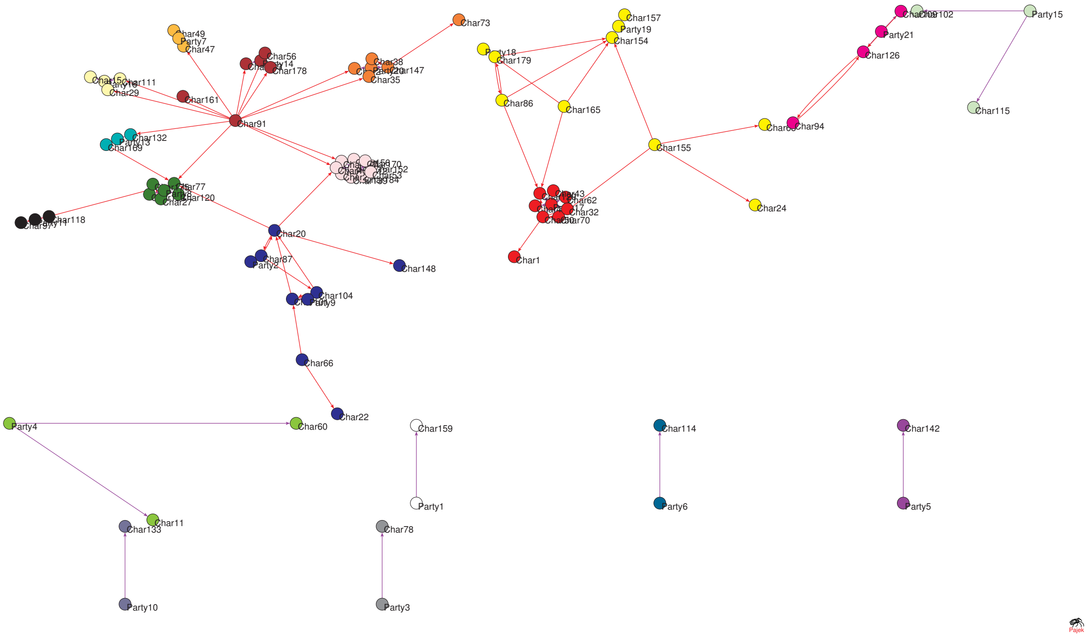

Figure 3.5: Clustering using Louvian community clustering method with default parameter values gives 19 clusters

# 4. Chat

Time references mentioned further are based on minutes elapsed from the time testing period started, i.d. a short time before first relevant record was saved (2015-6-1 13:00).

## 4.1 Whisper

Whispering is an action of sending a textual message from one player to another. Whispering is private, i.e. only sender and receiver can see the message which was sent. Analysis of whisper data shows which players communicate the most. Following data are results of the gathered whisper data being analysed in the whole, at the end of data gathering period.

Using Louvian clustering method, with default parameter values, seven communities were detected. These communities are shown in Fig. 4.1.

Many characters are strongly connected since they exchanged a lot of messages. Computed betweenness centrality detects three top nodes: Char77, Char175, Player. Their somewhat high centrality signifies that they stand in information flow between many characters, yet obviously more flows exist, since their betweenness is not significantly high.

Using R with network and ndtv libraries, heatmap was generated depicting whispering intensity of each pair of players, as shown in Fig. 4.3. It is observable that characters with high betweenness centrality have a lot of tracked traffic.

Combining data about party membership, social relations and whispering activity provides one with additional information on who the players communicated with, regarding their party membership. Network was generated using Union of Lines command in Pajek, combining network of social relations and party membership, and network of whispering activity. Nodes in Fig. 4.5 are coloured according to Strong P-Cliques partition with proportion of linkage with members of group equal to 0,9. Arcs representing party membership, social relations and whispering activity are coloured purple, red and black respectively.

Even simple analysis yields interesting information. Graphing collected data using R, some insightful graphs were constructed, as shown below.

In Fig. 4.6, where number of messages per character is shown, it is immediately visible that most of the whispering traffic was caused by only a couple of users. This information is consistent

| **Betweenness Centrality** | **Character** |
|---|---|
| 0,5687 | Char77 |
| 0,3230 | Char175 |
| 0,2281 | Player |
| 0,1280 | Char20 |
| 0,1044 | Char117 |
| 0,1041 | Char101 |
| 0,1024 | Char91 |
| 0,0555 | Whisper |
| 0,0532 | Char94 |
| 0,0526 | Char132 |
| 0,0526 | Char52 |
| 0,0526 | Char147 |
| 0,0508 | Char43 |
| 0,0463 | Char2 |
| 0,0200 | Char104 |
| 0,0116 | Char40 |
| 0,0092 | Char126 |
| 0,0057 | Char109 |
| 0,0014 | Char66 |
| 0,0008 | Char120 |
| 0,0003 | Char27 |
| 0 | Char47 |
| 0 | Char50 |
| 0 | Char32 |
| 0 | Char49 |
| 0 | Char111 |
| 0 | Char134 |
| 0 | Char87 |
| 0 | Char118 |
| 0 | Char35 |
| 0 | Char183 |
| 0 | Char57 |
| 0 | Char65 |
| 0 | Char139 |
| 0 | Char53 |
| 0 | Char146 |
| 0 | Char169 |
| 0 | Char29 |
| 0 | Char21 |

Table 4.1: Betweenness Centrality measures of nodes representing players who used whispering

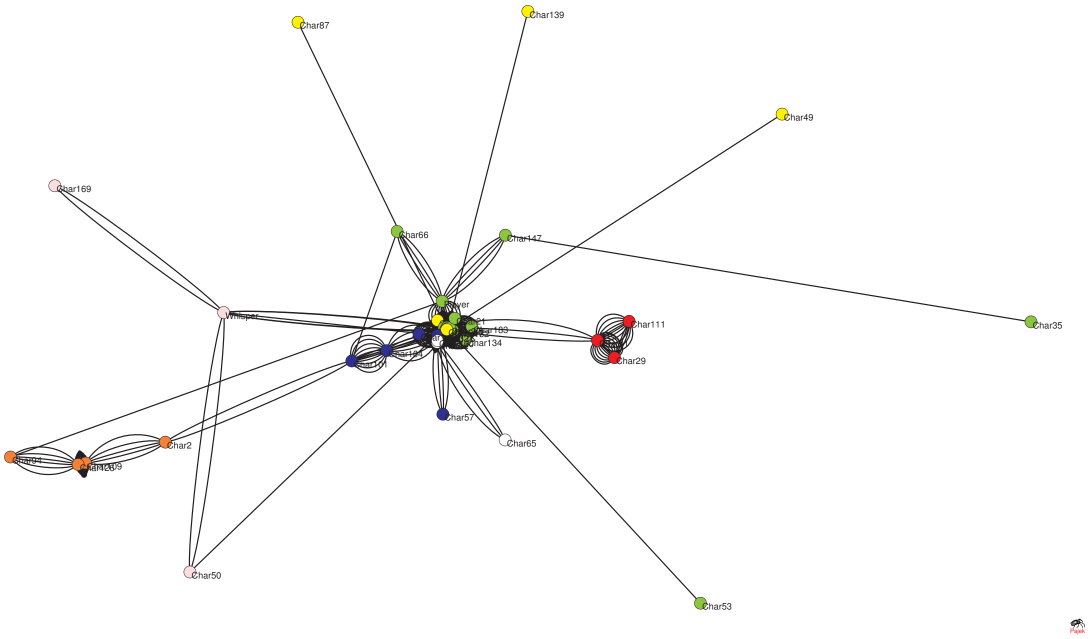


Figure 4.1: Clustering using Louvian community clustering method with default parameter values gives 19 clusters

with structure of networks shown earlier. Five most active characters, which account for 69,80% of whisper message traffic are shown in Tbl. 4.2.

| Character | # of messages |
|---|---|
| Char117 | 292 |
| Char40 | 424 |
| Char175 | 506 |
| Char77 | 757 |
| Char20 | 855 |

Table 4.2: Five characters with the greatest number of sent whisper messages

It is interesting to observe exchange of whisper messages through the whole time of testing period, and their fluctuations. Daily activity of users can be observed in Fig. 4.8, where most intensive whisper message exchange happens in regular intervals.

Data visualized in Fig. 4.8 represents how long a user was active in the game, in total. Time of the first recorded whisper message of a user is shown on x-axis, and coupled with time of the last recorded whisper message of a user (shown on y-axis) indicates total time interval of a user's activity in the game. Three of the longest active users are shown in Tbl. 4.3.

| Character | First Time | Last Time |
|---|---|---|
| Char77 | 9 | 31839 |
| Char175 | 16 | 31700 |
| Char20 | 1362 | 31659 |

Table 4.3: Three of the characters with some of the earliest and some of the latest recorded whisper time

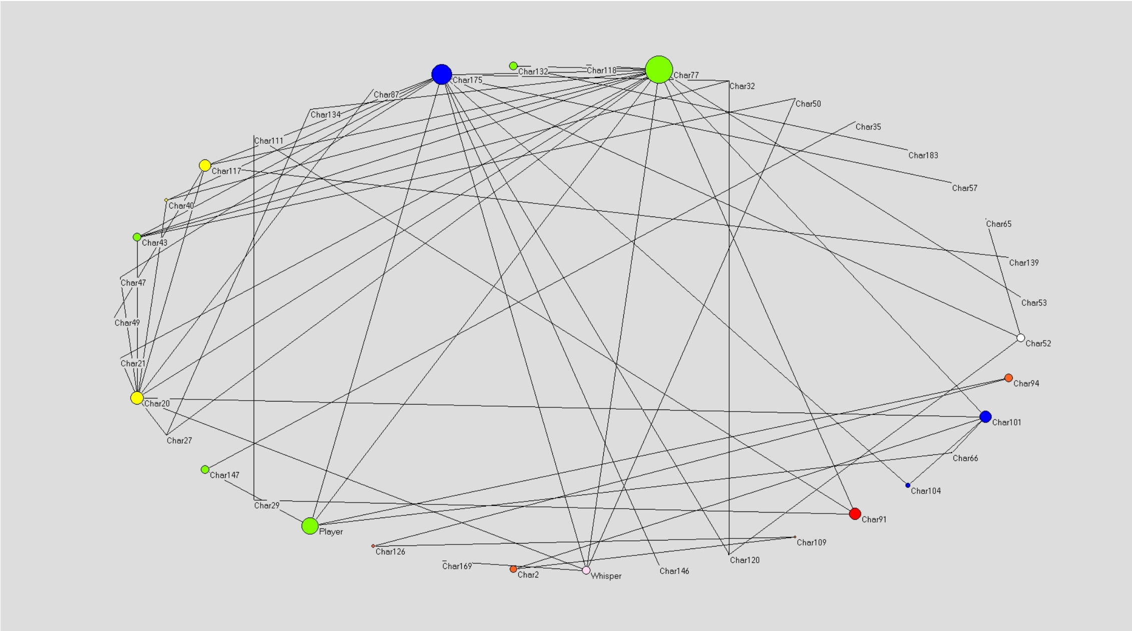


Figure 4.2: Louvian communities with node sizes relative to node betweenness centrality measure

## 4.2 Party Chat

Party messages are visible to party members only, and are a perfect medium to discuss, organize and cooperate a party. Usually containing content more public then that of a whisper message, but not public enough for general chat. As is visible in Fig. 4.9, only five users were significantly active in party message exchange. These five user's characters are named in Tbl. 4.4.

| Character | # of messages |
|---|---|
| Char49 | 3550 |
| Char77 | 1521 |
| Char175 | 1216 |
| Char134 | 1035 |
| Char27 | 663 |

Table 4.4: Most of the five characters which are most active in party chat are members of the same party

Party message exchange diminished in time, as is shown in Fig. 4.10. As the time passed by, parties got more passive and organization was not as crucial as before. It would seem players no longer needed information on where to find success in game, and time of interest passed, leaving in the game only those keen on adventuring.

Figure 4.11 confirms that most users used party chat for short periods of time only, and that most party chat activity stopped about time period 35000.

## 4.3 Trade Chat

When a user wants to trade some item for another item or in-game money, they must start a trade chat. Trading activity was sparse during the testing period, as visualized in Fig. 4.12. It is evident that trading was attractive at some moments only.


Figure 4.3: Heatmap showing whispering intensity for every pair of players; generated using R

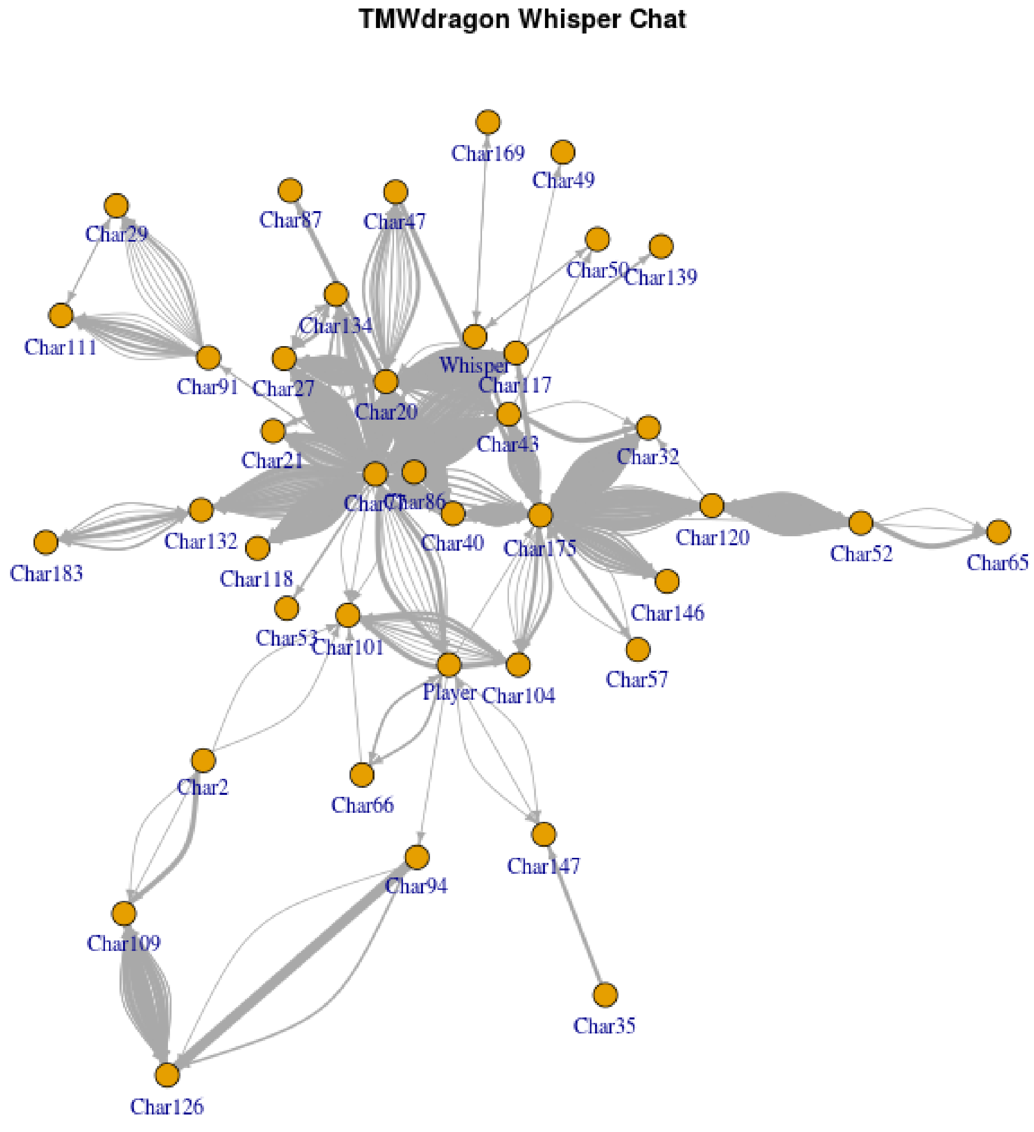


Figure 4.4: All whisper messages exchanged between users during the testing period may indicate that several users were main sources of information; generated using R

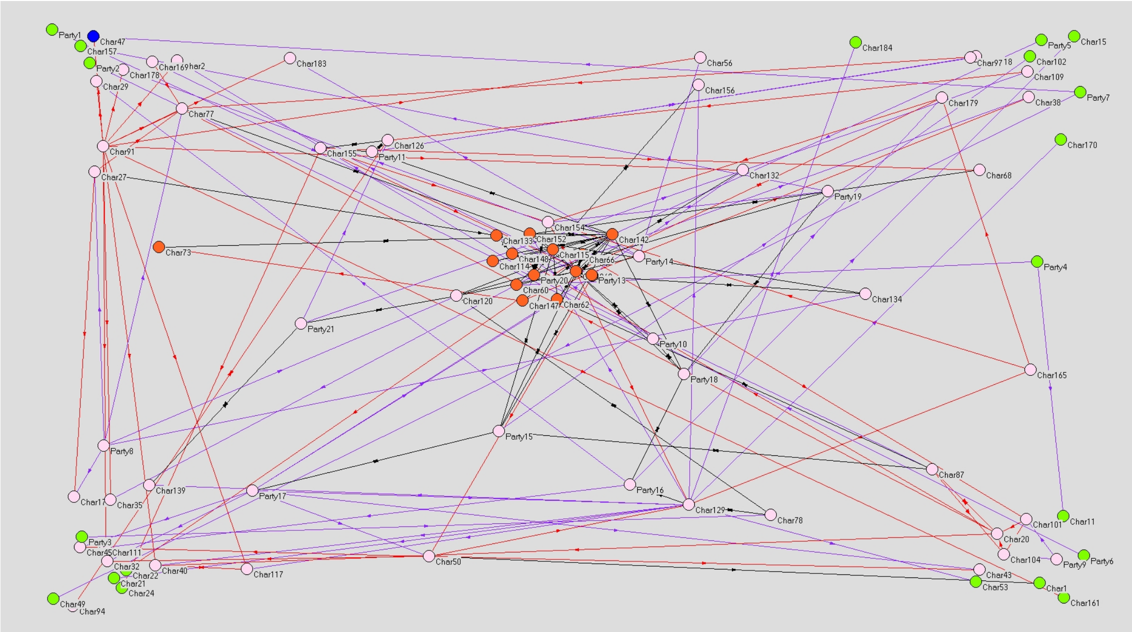

Figure 4.5: Heatmap showing whispering intensity for every pair of players

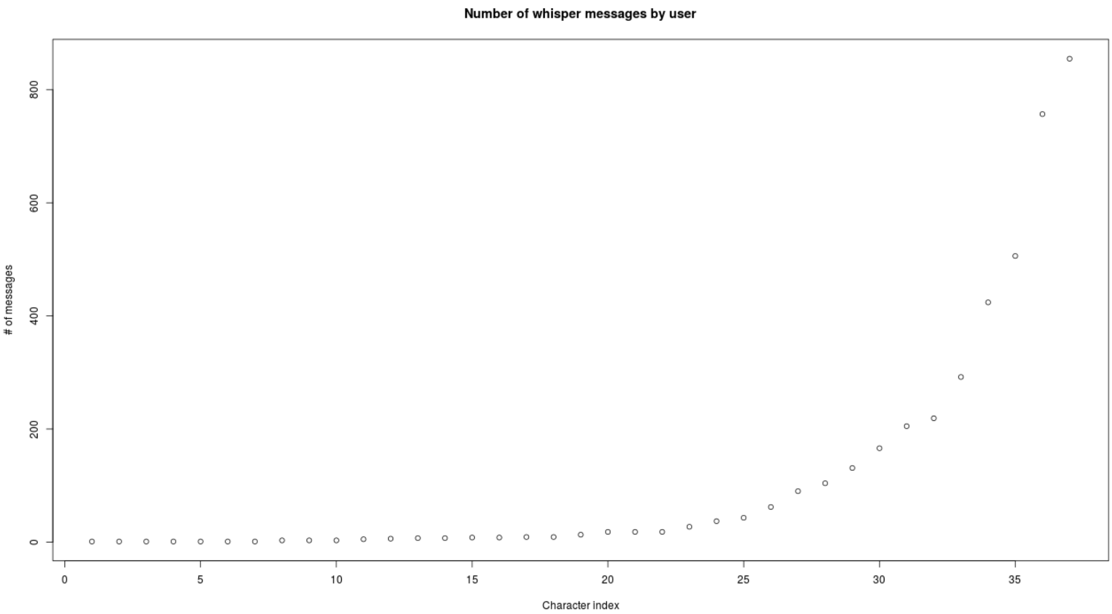


Figure 4.6: Number of whisper messages by character makes dominance of some users over whispering data evident

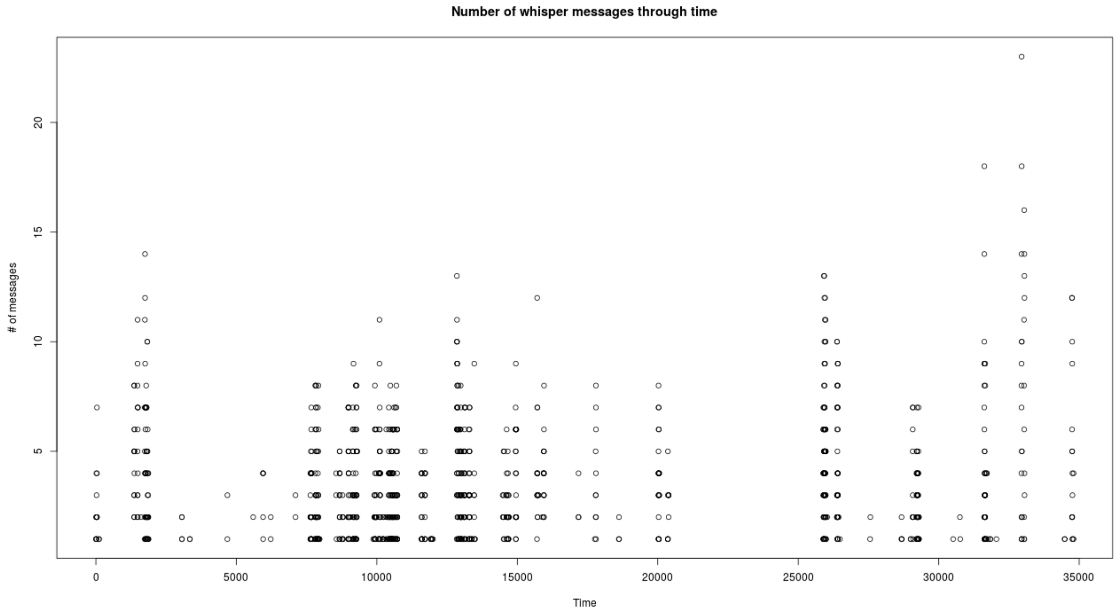


Figure 4.7: Number of whisper messages through time creates visual representation of when users have been most active in the game

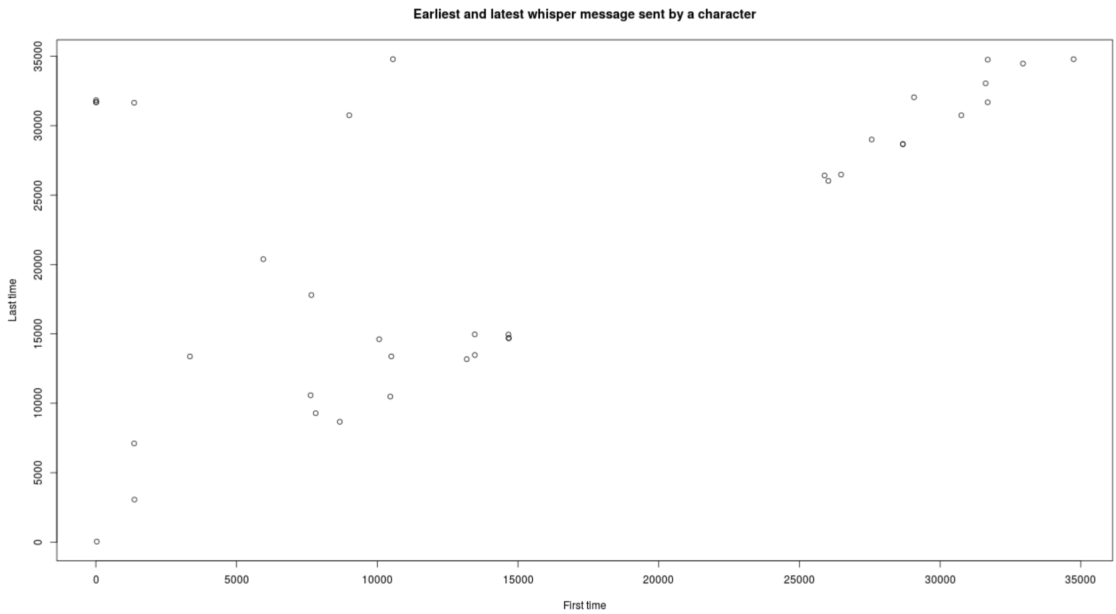


Figure 4.8: Time evidence of the first and the last whisper message of a user indicates for how long has the user been actively involved with the game

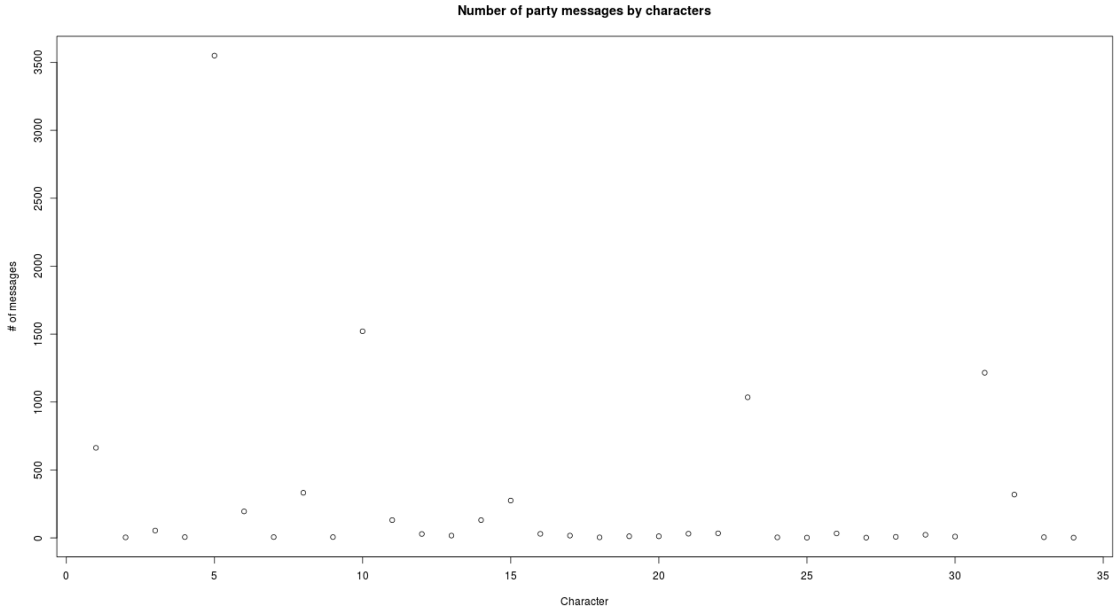


Figure 4.9: Significantly high number of party messages may indicate strong inter-party organization

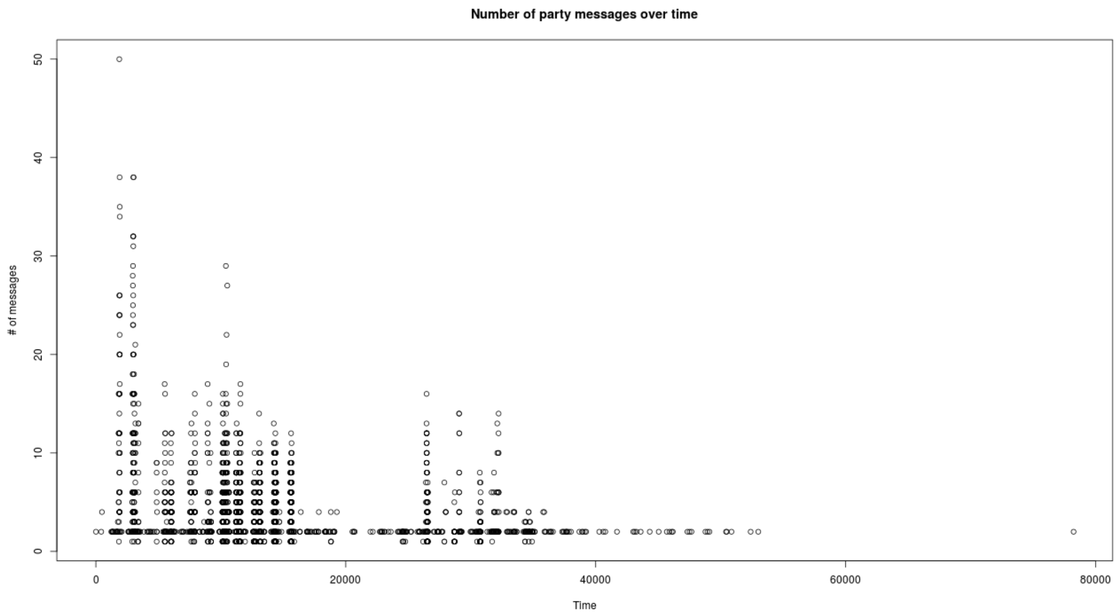


Figure 4.10: Where whisper chat ceased to exist (Fig. 4.7), party chat persisted, although with minimal activity

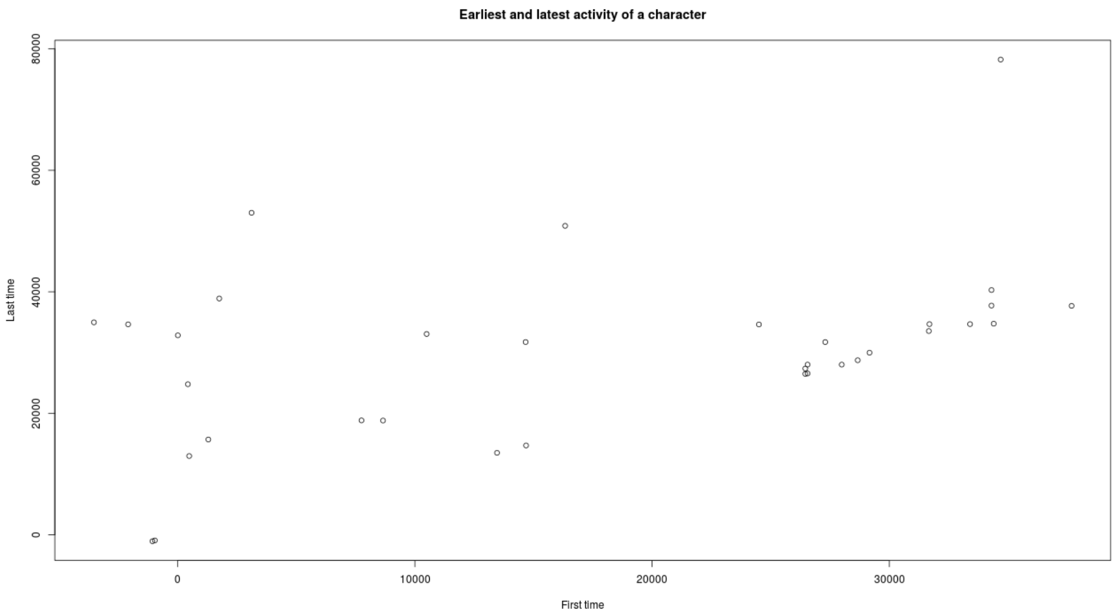


Figure 4.11: Only a few players stopped using party chat rather early in the game, indicating that players needed to cooperate and communicate in their chosen parties

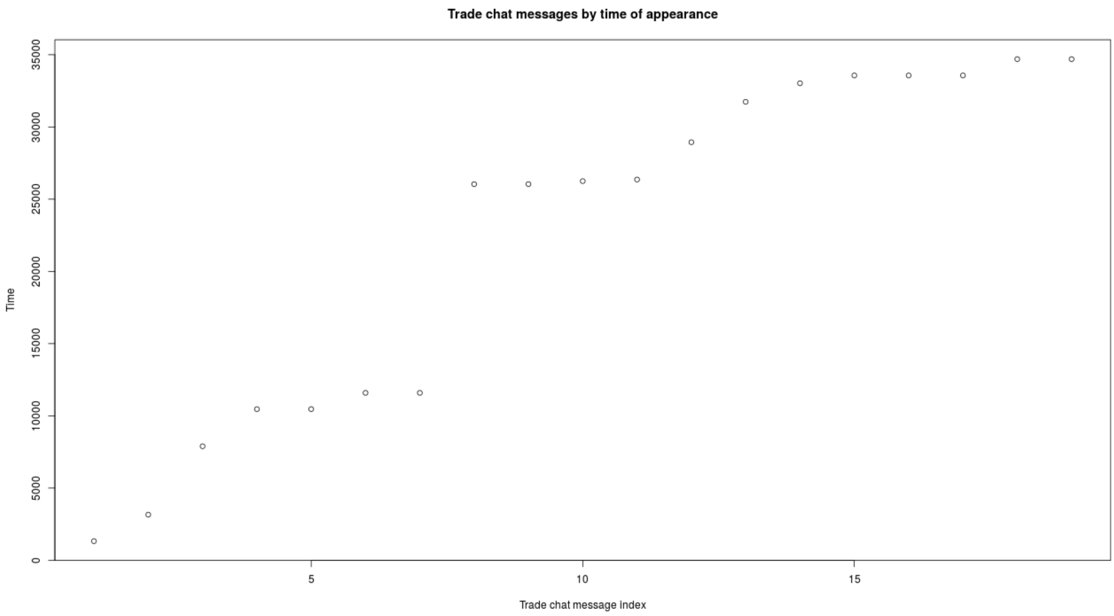


Figure 4.12: Visualized trade chat message time data may indicate that players used trade chat only when somebody reminded them of the possibility

# Bibliography

[1] Miroslav Bača, Markus Schatten, and Kornelije Rabuzin. "Towards an open biometrics ontology". In: *Journal of Information and Organizational Sciences* 31.1 (2007), pages 1–11 (cited on page 14).

[2] Mirjana Pejić Bach, Markus Schatten, and Zrinka Marušić. "Data Mining Applications in Tourism: A Keyword Analysis". In: *Central European Conference on Information and Intelligent Systems*. Faculty of Organization and Informatics Varazdin. 2013, pages 26–32 (cited on page 9).

[3] Kevin Bretonnel Cohen. "Natural Language Processing for Online Applications: Text Retrieval, Extraction and Categorization (review)". In: *Language* 80.1 (2004), pages 178–178 (cited on page 15).

[4] Magda Fontana and Pietro Terna. "From Agent-based models to network analysis (and return): the policy-making perspective". In: *SwarmFest*. University of Notre Dame. 2014 (cited on pages 13, 14).

[5] Goran Glavaš, Jan Šnajder, and Bojana Dalbelo Bašić. "Semi-supervised acquisition of Croatian sentiment lexicon". In: *Text, Speech and Dialogue*. Springer. 2012, pages 166–173 (cited on page 17).

[6] Global Collect. *THE GLOBAL MMO GAMES MARKET: Payments, Intelligence and Trends*. 2014. URL: `http://www.globalcollect.com/the-global-mmo-games-market` (cited on page 2).

[7] A. Kapetanović, M. Tadić, and D. Brozović-Rončević. "The Croatian language in the digital age / Hrvatski jezik u digitalnom dobu". In: *White paper series*. Springer, 2012, page 101 (cited on pages 16, 17).

[8] Niklas Luhmann. *Soziale Systeme: Grundriß einer allgemeinen Theorie*. Frankfurt, Germany: Suhrkamp, 1984 (cited on page 5).

[9] Peter Mika. “Ontologies are us: A unified model of social networks and semantics”. In: *Web Semantics: Science, Services and Agents on the World Wide Web* 5.1 (Mar. 2007), pages 5–15 (cited on pages 10, 11).

[10] Yamir Moreno, Maziar Nekovee, and Amalio F Pacheco. “Dynamics of rumor spreading in complex networks”. In: *Physical Review E* 69.6 (2004), page 066130 (cited on page 13).

[11] Martin Neumann, Armano Srbljinovic, and Markus Schatten. “" Trust Me, I Know What I'm Doing!" Competence Fields as a Means of Establishing Political Leadership”. In: *European Quarterly of Political Attitudes and Mentalities* 3.2 (2014), page 18 (cited on page 5).

[12] NewZoo. *The Global MMO Market; Sizing and Seizing Opportunities*. 2012. URL: `http://www.newzoo.com/infographics/the-global-mmo-market-sizing-and-seizing-opportunities/` (cited on page 2).

[13] Bo Pang and Lillian Lee. “Opinion mining and sentiment analysis”. In: *Foundations and trends in information retrieval* 2.1-2 (2008), pages 1–135 (cited on page 17).

[14] Romualdo Pastor-Satorras and Alessandro Vespignani. “Epidemic spreading in scale-free networks”. In: *Physical review letters* 86.14 (2001), page 3200 (cited on page 13).

[15] Gordan Ponjavić et al. “Personality Disorders in Croatia's Cyber Political Networks – A Topic Map Study”. In: *Proceedings of the 21st Central European Conference on Information and Intelligent Systems*. Edited by Boris Aurer, Miroslav Bača, and Markus Schatten. Varaždin: Faculty of Organization and Informatics, 2010, pages 33–39 (cited on page 10).

[16] Markus Schatten. “Mining Social and Semantic Network Data on the Web”. In: *Seminar za metodologijo in informatiko*. Faculty of Information Studies, 2011 (cited on page 5).

[17] Markus Schatten. “Opinion Mining on News Portal Comments – Towards a Croatian Sentiment Database”. In: *Seminar za metodologijo in informatiko*. Novo Mesto: Fakulteta za informacijske študije, 2012 (cited on page 17).

[18] Markus Schatten. “What do Croatian Scientist Write about? A Social and Conceptual Network Analysis of the Croatian Scientific Bibliography”. In: *Interdisciplinary description of complex systems* 11.2 (2013), pages 190–208 (cited on pages 10–12).

[19] Markus Schatten, Jurica Ševa, and Bogdan Okreša Đurić. “Big Data Analytics and the Social Web-A Tutorial for the Social Scientist”. In: *European Quarterly of Political Attitudes and Mentalities (EQPAM)* 4.3 (2015), pages 30–81.

[20] Markus Schatten, Jurica Ševa, and Okreša Đurić, Bogdan. “An Introduction to Social Semantic Web Mining & Big Data Analytics for Political Attitudes and Mentalities Research”. In: *European Quarterly of Political Attitudes and Mentalities EQPAM* 4.1 (2015), pages 40–62 (cited on pages 2, 5).

[21] Markus Schatten et al. “An Analysis of the Social and Conceptual Networks of CECIIS 2005 – 2010”. In: *Proceedings of the 22nd Central European Conference on Information and Intelligent Systems*. Edited by Tihomir Hunjak, Sandra Lovrenčić, and Igor Tomičić. Varaždin: Faculty of Organization and Informatics, 2011, pages 259–264. (Cited on pages 8, 10).

[22] G. Smith. *Tagging: People-Powered Metadata for the Social Web*. New Riders, Berkeley, 2008 (cited on page 9).

[23] IHsien Ting. “Web mining techniques for on-line social networks analysis”. In: *Service Systems and Service Management, 2008 International Conference on*. IEEE. 2008, pages 1–5 (cited on page 6).

[24] Camelia Florela Voinea and Markus Schatten. “Recovering the Past. Eastern European Web Mining Platforms for Reconstructing Political Attitudes”. In: *European Quarterly on Political Attitudes and Mentalities* 4 (2014), pages 22–39 (cited on page 5).

[25] Wikipedia Contributors. *Massively multiplayer online game — Wikipedia, The Free Encyclopedia*. [Online; accessed 9-February-2015]. 2015. URL: `http://en.wikipedia.org/w/index.php?title=Massively_multiplayer_online_game&oldid=646153954` (cited on page 2).

[26] Wikipedia Contributors. *Massively multiplayer online role-playing game — Wikipedia, The Free Encyclopedia*. [Online; accessed 9-February-2015]. 2015. URL: `http://en.wikipedia.org/w/index.php?title=Massively_multiplayer_online_role-playing_game&oldid=642752380` (cited on page 2).

[27] Wikipedia Contributors. *Role-playing video game — Wikipedia, The Free Encyclopedia*. [Online; accessed 9-February-2015]. 2015. URL: `http://en.wikipedia.org/w/index.php?title=Role-playing_video_game&oldid=642919862` (cited on page 2).

[28] Guandong Xu, Yanchun Zhang, and Lin Li. *Web mining and social networking: techniques and applications*. Volume 6. Springer Science & Business Media, 2010 (cited on page 6).

[29] Jianmei Yang et al. “A study of the spreading scheme for viral marketing based on a complex network model”. In: *Physica A: Statistical Mechanics and its Applications* 389.4 (2010), pages 859–870 (cited on page 13).