DESIGNING ALGORITHMS FOR REALISING SOCIAL GOALS

by
Radosław Nielek

Supervisor
Professor Adam Wierzbicki

A dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy (Computer Science)
at the
Polish-Japanese Institute of Information Technology

2012

Science advances whenever we can take something
that was once invisible and make it visible;
and this is now taking place with regards
to social networks and social processes.

Jon Kleinberg

Acknowledgement

I would like to express my deep and sincere gratitude to my supervisor, Professor Adam Wierzbicki, for his comments, hours of discussion and unwavering belief that I will finish this dissertation in a finite time. His passion for teaching and ability to get things done were constantly motivating and inspiring at the same time.

I am particularly grateful to Professor Romuald Kotowski, who drew my attention to scientific questions and helped me to write my first scientific paper.

Efficient work is teamwork, thus I am indebted to people working with me during the time I was writing my dissertation. I warmly thank Aleksander Wawer, Tomasz Kaszuba and Piotr Turek for the research and publications we have jointly done.

I wish to thank my parents for their encouragement and constant support.

I am also indebted to the people who helped me to prepare this document. Kinga and Piotr, thank you for your assistance.

Table of Contents

Streszczenie

Praca „Projektowanie algorytmów wspierających realizację celów społecznych" koncentruje się na poszukiwaniu odpowiedzi na pytania czy algorytmy mogą wpływać na zachowania ludzi, a jeśli tak, to czy projektowanie takich algorytmów dla osiągnięcia założonego celu jest możliwe (i jak to zrobić). Praca składa się z czterech rozdziałów, z których pierwszy opisuje dotychczasowy stan wiedzy, przedstawia związek między systemami społecznymi a informatycznymi, a także definiuje tezy badawcze. Drugi i trzeci prezentują oryginalne wyniki badań, a czwarty rozdział zawiera podsumowanie i opis możliwych kierunków dalszych badań. Praca kończy się spisem literatury z ponad 150 pozycjami, listą rycin i tabel. Celem rozprawy było dowiedzenie dwóch tez postawionych we wstępie:

– **Teza I:** Konstrukcja systemu informacyjnego może pozytywnie wpływać na zachowania użytkowników i pojawiające się w nim zjawiska społeczne.
– **Teza II:** Można projektować algorytmy w taki sposób, aby wspierały osiąganie założonych z góry celów społecznych.

Pierwszy rozdział rozpoczyna się od wstępu teoretycznego poświęconego związkom między systemami społecznymi i informacyjnymi, a także badaniom prowadzonym nad wpływem jednych na drugie. Kolejne punkty definiują pojęcie celów społecznych rozumianych w ramach tej pracy jako powszechnie akceptowane, uzgodnione w sposób świadomy lub nieświadomy cele danej społeczności a także prezentują kluczowe pojęcia i narzędzia dla projektowania algorytmów wspomagających realizację celów społecznych, czyli zbierania danych behawioralnych, emergencji, automatycznego przetwarzania języka naturalnego i symulacji społecznych. Rozdział dzieli główny problem na dwa komplementarne problemy (zagadnienia), których rozwiązanie potwierdza postawione w pracy tezy:

– analiza zjawisk społecznych na podstawie danych behawioralnych z systemów teleinformatycznych,
– celowe modyfikowanie zjawisk społecznych przez odpowiednią konstrukcję i modyfikację algorytmów.

W rozdziale drugim pokazano, że identyfikacja zjawisk społecznych w danych uzyskanych z systemów teleinformatycznych jest możliwa. Badania przeprowadzono na dwóch różnych typach danych i dotyczyły zjawisk pojawiających się zarówno w ramach dość specjalistycznego systemu (spirala nienawiści na portalu aukcyjnym) jak i bardziej uniwersalnych, dotyczących de facto całego społeczeństwa (predykcja cen akcji na giełdzie w Warszawie). Pierwsza część rozdziału poświęcona jest analizie zbioru danych dotyczącego zachowań użytkowników największego polskiego serwisu aukcyjnego – Allegro. Analizy statyczne pozwoliły znaleźć ciekawe prawidłowości w zachowaniach zarówno sprzedawców jak i kupujących (m.in. kolejność umieszczania komentarzy czy długość komentarzy). Te analizy zostały uzupełnione o dokładniejsze studia nad zawartością komentarzy wykonane przy pomocy algorytmów do automatycznego

przetwarzania języka naturalnego. Do najciekawszych, otrzymanych w tej części pracy wyników można zaliczyć: wykazanie dużego podobieństwa między komentarzami neutralnymi i negatywnymi, identyfikację zjawiska amplifikacji agresji w komentarzach wśród osób, które wystawiają je jako drugie (po przeczytaniu oceny jaka im została wystawiona), a także generalną niską informatywność w treściach komentarzy (zbliżającą się do zera dla komentarzy pozytywnych). Dodatkowo, zaproponowana została modyfikacja algorytmu oceniania, która pozwala wyeliminować większość zidentyfikowanych negatywnych zjawisk. W drugiej części rozdziału pokazano, że nawet dla tak złożonych zjawisk jak predykcja cen akcji na giełdzie możliwe jest znalezienie pewnych zależności między danymi zbieranymi w Internecie i analizowanymi za pomocą dedykowanych algorytmów przetwarzania języka a samym zjawiskiem.

Przykłady konstrukcji algorytmów wspierających realizację celów społecznych zostały przedstawione w rozdziale trzecim. Na przykładzie systemu zarządzania zaufaniem wykazano możliwość tworzenia takich algorytmów. Kolejno zostały omówione i wykonane wszystkie kroki niezbędne przy projektowaniu. Cel sprawiedliwej dystrybucji dóbr został sformalizowany w sposób umożliwiający jego mierzenie i porównywanie. W celu eksperymentalnego sprawdzenia proponowanych algorytmów stworzono symulator systemu aukcyjnego, który umożliwiał testowanie odporności proponowanych algorytmów dla różnych strategii adwersarzy. W celu zwiększenia realizmu symulacji, a także uwzględnienia faktu, że rzeczywiste system aukcyjne są otwarte, dane z rzeczywistego systemu aukcyjnego zostały użyte do sterowania zachowaniem części agentów. Zaproponowane algorytmy okazały się efektywne w realizacji wyznaczonego celu przy okazji nie zmniejszając liczby transakcji w serwisie aukcyjnym, co było kluczowym elementem determinującym ich potencjalne, praktyczne zastosowanie.

System reputacyjny w serwisach aukcyjnych jest scentralizowany stąd konstruowanie dla niego algorytmów realizujących cele społeczne ma pewną specyfikę. W celu wykazania, że takie algorytmy mogą powstawać także dla rozproszonych systemów w dalszej części trzeciego rozdziału zaproponowano rozwiązanie, które umożliwia analizę emocji użytkowników telefonów komórkowych, a także wpływanie na nie przez filtrowanie informacji nacechowanych emocjonalnie, odpowiednio pozytywnie lub negatywnie.

W toku badań udało się wykazać, że dla trzech typów danych pozyskanych z różnych źródeł, począwszy od urządzeń mobilnych użytkowników, poprzez strony internetowe, a skończywszy możliwa jest identyfikacja zjawisk społecznych. Dodatkowo, zaprojektowanie algorytmów wspierających realizację celów społecznych zarówno dla scentralizowanego systemu reputacyjnego, jak i rozproszonego świata użytkowników smartfonów pokazuje, że cele takie mogą być stawiane przed twórcami algorytmów. Dalsze badania powinny być prowadzone w kierunku rozbudowy mechanizmów umożliwiających wyrażanie celów społecznych w sposób sformalizowany, ich mierzenie i porównywanie. Znaczącym ułatwieniem dla rozwoju i praktycznego zastosowania tej dziedziny wiedzy byłoby także stworzenie uniwersalnych narzędzi symulacyjnych pozwalających, po tylko niewielkich zmianach, testować propozycje nowych algorytmów dla szerokiej klasy systemów.

1. Introduction

1.1. Overview and contributions

Can algorithms influence people's behaviours and if yes, can we design algorithms which realise social goals? These two questions will be the central issue of this dissertation. A broad range of algorithms dedicated to solving well-defined issues, like sorting or traveling salesman problem, deliver results regardless of the social context. Additionally, their performance can be evaluated with objective criteria like computational complexity or correctness. Ordering products by price is an easy and measureable task, ordering by relevance to users' expectations requires a much more complicated algorithm and, which is even more challenging, a good and computable definition of user expectations. Even more troublesome are algorithms working in a rich, labile social context like trust management systems in auction houses. Outcomes delivered by such algorithms cannot be meaningfully interpreted outside the information system and, even then, a subjectivity problem has to be addressed – what is a good deal for the seller at the same time can be fraud for the buyer. Constantly changing strategies of social actors require also a flexible adaptation mechanism embedded into the algorithm.

Algorithms, next to interaction between people, data and processes, are an indispensable part of information systems. Therefore, in all socially-centric technologies, which are dedicated to satisfy some people's expectations and needs, part of responsibility has to be transferred to algorithm design process. Taking into account additional requirements, algorithms originators must apply a different approach and address many unexpected issues. On the one hand, it is impossible to verify algorithms outside information system; on the other hand, it is hard to imagine that each version will be tested in the real productive system, e.g. online auction house or social network site. This vicious circle can be broken either by abandoning the try-and-fix approach or by social simulation described in detail in chapter 1.4.

Social simulation is powerful enough to verify the most promising solution but even a very extensive simulation does not guarantee that the outcome in the real system will be the same. Despite very good models of physical processes and powerful computers, the certificate for an aircraft is granted based on the performance of the manufactured aircraft and not its virtual model. We do not have models of social processes, which are even closer in a matter of precision to physical ones, so the final tuning has to be based on the performance of an algorithm in a real-world implementation, in the rich context delivered by the information system. Going out of the laboratory to the real world complicates the process of collecting data necessary to conduct evaluation. Privacy concerns, incomplete information and the need for selecting *a priori* relevant data are only the beginning of problems encountered during data collection. More about such problems and possible solutions is written in chapter 1.3.

Even the most complete dataset about interaction in information system does not solve the very fundamental problem – how to measure the realisation of the social goal and what social goals mean in reality. Skipping for a while a precise definition, which will be discussed more indepthly in chapter 1.2.2., for now let us mention only a few possible social goals: justice, fairness, equity, peace, wellbeing, agreement or security. There is a common agreement among society members that goals like listed above are desirable but at the same time there are fierce arguments what these terms mean and what is the best method to realise them. The discussion about the definition of justice or wellbeing and theoretical and practical consequences of various propositions are far beyond the scope of this thesis but one detail is important for algorithm designers – how the level of realisation of social goals can be computed. In chapter 3.1. some methods are shown for computing the level of fairness in information systems and a similar process is required for every social goal, which has to be realised by an algorithm.

The most sophisticated tools and approaches used by algorithms developers will not matter if the algorithms and technology do not affect people's behaviours. A common sense assumption that between social and information system exists a bidirectional feedback is not taken for granted by sociologists. In a hot discussion sociologists support very different views – from the claim that the development of technology is predetermined in advance to the more balanced view that social actors decide about the technology used and its applications. More about the theoretical foundation of connection between social science and informatics can be found in chapter 1.2.1.

The scope of this dissertation is, in the first step, to verify the assumption that the design of an algorithm and information system can influence users' behaviours and that this impact can be, to some extent, predicted. In addition, if the interconnection stated in the previous sentence exists, then it is possible to design algorithms which support reaching predefined social goals. To present the problems mentioned above more formally, two hypotheses have been defined:

– **Thesis I:** The design of information system can positively influence people's behaviours and social phenomena,
– **Thesis II:** It is possible to construct algorithms that help achieve predefined social goals.

This dissertation contributes to the current state of knowledge in the following mutually interconnected areas: collecting data for sensing social phenomena, processing unstructured data, formal definitions and measuring of social goals, designing algorithms which support the realisation of social goals, and verifying algorithms with the help of social simulations. In particular, some key achievements are listed below.

- It was shown that the relevant data containing information about social phenomena could be collected from at least three different sources with substantially different approaches. Among them:
  - centrally collected dataset from a production system which is focused on performance and delivering services rather than doing research (see chapter 2.1.),
  - collection of activity trails (i.e. comments, articles, blogs) dispersed in a heterogeneous system (see chapter 2.1. and 2.2.),
  - individually focused dataset of emotional states of users reconstructed on the basis of behavioural data collected from smartphones (see chapter 3.2.).

  Additionally, as some datasets contain Polish and some English texts, a universality of the proposed approach was proved.
- A new, multi-criteria function was proposed for a simultaneous evaluation of an equitable distribution of goods at auction houses and the aggregate number of transactions (see chapter 3.1.). This function takes into account not only buyers' and sellers' objectives but also the interests of the owner of an auction site, and is a good example of a formal definition of social goals.
- The existence of a spiral of hatred phenomenon in currently widely used reputation systems has been discovered and its influence on the quality of reputation systems has been studied. Additionally, some countermeasures have been proposed (see chapter 2.1.).
- The feasibility of designing algorithms for supporting the realisation of social goals was confirmed by the development of such algorithms. In particular, the following algorithms have been proposed:
  - reputation algorithm which support fair distribution of payoffs in auction houses and is resistant to a broad class of adversaries' strategies,
  - application for smartphones which senses people's mood and tries to influence it by either filtering out some information or by choosing an appropriate wallpaper (see chapter 3.2.),
  - algorithm for reputation evaluation which will eventually increase the trustworthiness of reviews of transaction outcomes and limits the unnecessary stress connected with negative evaluation (see chapter 2.1.).
- Trace-driven social simulation was proposed and positively tested as an invaluable tool for verifying and developing socially aware algorithms.
- Trace-driven social simulation was proposed and positively tested as a valuable tool for verifying and developing social-aware algorithms.

The remainder of the dissertation is organized as follows. In the first chapter theoretical and methodological issues are discussed, among them problems with data collections and terms of using social simulation as a scientific tool. The second chapter is focused on methods for sensing social phenomena in information systems, and two examples are studied more indepthly. The first subsection is devoted to the interconnection between mood and market. The second subsection presents social phenomena existing in reputation system and explains how they are influenced by construction of information systems and, in particular, by algorithms. The third chapter describes methods of designing algorithms shaping social phenomena. The first subsection in the third chapter shows by going through all steps – from formal description of requirements to testing scenarios by social simulation – how to create a new reputation algorithm, which assures equity to users. The last subsection in the third chapter presents a mobile application which allows influencing people's mood either by filtering good/bad news or by delivering special crafted messages. The fourth chapter briefly summarizes the research presented in this dissertation and mentions the promising direction for future investigations.

1.2. Social and Information systems

1.2.1. From social to information system

The increasing convergence between social and information systems has been firstly noticed more than three decades ago (Weizenbaum 1976; Mowshowitz 1976). Researchers into this phenomenon came from various disciplines – from computer science and management to sociology, psychology and even philosophy. Different backgrounds caused a never-ending dispute about the scope of social informatics (SI) and its definition. Dr Rob Kling who is commonly believed to be a father[1] of the term "Social Informatics" defined this discipline as "the interdisciplinary study of the design, uses and consequences of information technologies that takes into account their interaction with institutional and cultural contexts" (Kling 1999). Institutional and cultural context appeared in the definition because at that time various organizations were early adopters of information technologies, and was later replaced with a more general term – "social context". Still, such a definition introduces a kind of hierarchy where the main area is technology and the social context plays only a complimentary role. A more balanced definition was proposed by Lamb and Sawyer and states that SI is "a body of rigorous empirical research that focuses on the relationships among people, ICTs, and the social structures of their development and use" (Lamb, Sawyer 2005).

Despite almost thirty years of its history, social informatics is still not widely accepted as an umbrella name for a discipline of science. On the one hand, many

[1] Vasja Vehovar, a Slovenian scientist, claims that the very first use of the term "social informatics" was in 1985 at the Faculty of Social Science of the University of Ljubljana to label the four-year undergraduate program (Vehovar 2006) but since they used only the Slovenian term – *"Dru boslovna informatika"* – the authorship of the English one is still attributed to Rob Kling. Around the same time the Norwegian Ministry of Education established SI as a discipline at the University of Oslo (Grosseck 1985).

well-established universities, among them Kyoto University, Purdue University or Oxford Internet Institute, have opened either departments, studies or courses in social informatics[2] but on the other hand many researchers are attempting to popularise alternative terms like "internet sociology", "e-society science" or "social simulation" for a very similar, or even the same, scope of research. Particularly interesting is situation in Japan where "social informatics" appears next to the term "socio-informatics" that seems to be slightly more popular there, and each of them has its own association (Kurosu 2010). The huge overlap existing between these two terms has been noticed by Japanese scientists and, thus, JASI[3] and JSIS[4] publish a joint journal – The Journal of Socio-Informatics[5].

An interesting area of convergence between social science and informatics is social network analysis. Jacob L. Moreno conducted one of the first research projects dealing with social networks in 1930s, even before the first computer was built. A purely social concept was broadly assimilated and adapted by computer specialists mostly due to an unimaginable development of social websites, and pioneering works of Albert-László Barabási explain the emergence of scale-free networks in a variety of social and information systems. Yet other scientists active in areas like human-computer interaction, computational social choice or web mining mostly agree that social aspects of information technology are crucial to fully understand the appearing interaction but at the same time see their disciplines as separate research areas, at best loosely connected with social informatics. On the other hand, many researchers originating from computer science disciplines like computer networks or databases have accepted the term *social informatics* and stress its maturity and "unique set of research objectives and methods" (Bolc 2010).

The abrupt increase of Internet penetration, ubiquity of PCs and now also smartphones created a huge and novel system inextricably connecting society and technology. This process has a broad range of consequences reaching far beyond discussions about the social informatics definition. Jon Kleinberg draws attention to the fact that the ubiquity of computer devices in general and popularity of social media in particular creates a huge poll of behavioural data with minute-to-minute granularity (Kleinberg 2008). Another group of well-established researchers, among them Tim Berners-Lee and Nigel Shadbolt, stress that "the Web is the most used and one of the most transformative applications in the history of computing, even of human communication" (Berners-Lee, Hall, Hendler et al. 2006) and thus put in their effort into promoting a new science discipline – web science – which encompasses everything which is connected with the Internet and its applications (Berners-Lee, Hall, Hendler et al. 2006).

[2] An extensive lists of courses, studies and departments interconnected with social informatics can be found on http://www.social-informatics.org/c/151/Study_programs/ (03.2012).

[3] Japan Association for Social Informatics (http://wwwsoc.nii.ac.jp/jasi/).

[4] Japan Society for Socio-Informatics Studies (http://www.soc.nii.ac.jp/jsis/).

[5] http://wwwsoc.nii.ac.jp/jasi/eng/eng02.html.

Don Tapscot and Anthony D. Williams present a slightly different view. Authors of the well-received bestseller "Wikinomics: How Mass Collaboration Changes Everything" contend that the most important aspect of the Internet revolution is the opportunity to easily merge people's activities and knowledge and direct them to common goals (Tapscott, Williams 2006). In their next book they go even further and expect that crowdsourcing is the only way to "reboot business and the world" in the context of global crisis (Tapscott, Williams 2010).

Parallel to the proponents of the web science who identify the Internet as the central point of the new research movement, and worshippers of crowdsourcing, an active community supports another approach focusing on research into social phenomena with the help of social simulation. Robert Axelrod even claims that social simulation[6] is the third way of doing science, next to deductive and inductive approach (Axelrod 1998). Although Axelrod's statement seems to go too far, social phenomena as the main element connecting various disciplines mentioned above are very promising.

People bring technology to life and their interactions make the Web a really interesting place from the scientific perspective. Putting aside for a moment issues of scalability, the most famous websites like Twitter, Facebook, Digg or eBay are, from the technological point of view, not very challenging and, thus, not very interesting for scientists. What make such services extremely exciting are people's interactions that are intermediated by technology. The existence of an intermediary creates an opportunity to collect all details about people's behaviours with almost infinite resolution but attempts to understand social processes cannot be elided from the technological context. Even more, technology is not merely an intermediary or context. It is an inherent part of the system. The website design strongly influences how a number of web pages visited by user (Saraschandra 2010) placed in the Google results translates into the number of visitors (Bar-Yossef, Gurevich 2009); the hard-coded 140 characters limit[7] on the message length in Twitter makes not only URL shorteners boom (Demetris et al. 2011) but also modifies the language used by users (Crystal 2011; Gouws 2011).

Jon Kleinberg's postulated convergence of social and technological networks observable on social network sites (e.g. Facebook, GoldenLine) is *de facto* a new process of building such networks. Assuming an access to the Facebook data, the social network of users can be easily reconstructed and analysed but a closer look at

---

6 For scientists active in social simulation community the social phenomenon is usually inseparably connected with the term *emergence*. The over two-thousand-year-old dispute about the meaning of this term and philosophical consequences of existence or non-existence of such a process is far beyond the scope of this dissertation. An inspiring discussion can be found in (Sawyer 2005) and (Bedau, Humphreys 2007). The author of this dissertation stands on the side of reductionists.

7 140 characters limit on Twitter messages derived from the maximum length of the SMS in GSM networks. Twitter founders wanted to avoid the need to split one message for many SMSs and set the limit in their system on the same value but reserved 20 characters for user identification (Milian 2009).

the obtained network will probably raise a lot of questions. Does it really resemble either acquaintance or friendship network? Who is missing and why? One of the most important forces that shape the Facebook social network is the recommendation system which suggests who you may know[8]. The exact algorithms are not disclosed but some tries show that the recommendation is based not only the number of common friends but also types of friends and search history. The current solution leads to heavy "overfriending"[9] but also to some new relationships in the real world. On the other hand, even small changes in the algorithm will definitely influence the friendship network, both real and virtual.

The effect of bidirectional relations between information and social system presented in the previous paragraph is not a unique feature of either Facebook or social network sites in general. A careful look around reveals that almost every slightly more complicated website has similar mechanisms. Every Amazon customer knows the "Customers Who Bought This Item Also Bought" recommendation; the Gmail suggests users whom they might want to add to the email recipients; stars on Allegro decide which seller will be most successful. It is only a handful of examples. Yet another interdependency between social and information systems is presented in figure 1.1. Google supports visitors typing their queries by a list of hints. Every subsequent character is used to tune suggestions. No one but Google has data indicating how strong this mechanism influences what people are looking for but we can easily assume that visitors use hits just because the Google company still keep it as a part of their website. Like with the Facebook recommendation mechanism, Google do not reveal details about the suggestion mechanism but reverse engineering helps to confirm that it is quite a complicated solution[10]. Google takes into account not only the popularity of searched terms but also their co-occurrence and even personal search history and localisation.

Interactions between social and information systems drew the attention of sociologists who raise the question about causality in this relation. A variety of views can be classified into three main movements: substantivism, instrumentalism and constructivism. Instrumentalists see technology as a neutral instrument that cannot be judged in terms of good or bad and attribute the main developing drive to public and private investments, which in turn increase the importance of technology

8 As a curiosity, Facebook does not have a mechanism encouraging people to remove the edges. Therefore, the link once created never expires spontaneously, even if it is not used.

9 "Overfriending" is a by-product of the original assumptions about Facebook as a website, which helped Harvard students to keep in touch and know each other better (Kirkpatrick 2010). "I probably know a friend of my friend" heuristic worked very well for the closed community of Harvard students and not so well for the general public. Dense social network at the early stage of web site development improves the user experience by making the stream of news and events more dynamic but now is identified as one of the factors that limit growth.

10 Some information has been revealed due to the legal action taken by the French company against Google. Insurance company Lyonnaise de Garantie was offended by the Google suggestions mechanism which proposed to complete the search term containing company name with the French word "escroc" which means in English "crook" (Brodkin 2012).

in society. Substantivists assume that technology cause social changes which lead to the *a priori* defined state. They believe that technology in general, and the Internet in particular, tend to increase power and hegemony and such a process is inevitable. The main assumption of the third theory – constructivism – is that *"new technological systems emerge through a process of negotiation (…) involving a myriad of social actors"* (Khademhoseiny 2010). The term *social actors* should be understood very broadly as institution, organization or even, a group of people and the only condition is that they share the same set of meanings and goals. Social-centric technology is a social actor itself. A good example of this is the situation when Google has blocked certain Polish price-comparison sites because they were using search-engine spamming to appear high in the Google ranking. This move has a real influence on the Polish e-commerce market. Negotiations[11] between relevant social actors, understood as a by-product of varied forces and objectives, determine the development path for technology. To be more precise, the path is selected from a set of possibilities limited by the current state of technology.

Fig. 1.1. Google suggestions. Google is trying to guess the right query during typing and shows suggestion based on many factors (e.g. user location, and link popularity)

Socially centric platforms are social actors. They attempt to realize the social goals of entire communities. They can use social concepts, like trust, to better understand a social environment, and to better motivate users. Also, *a socially centric platform is not just a technology: rather, it is also a set of protocols that define the social practices and interactions of its users who are members of a community*. The goal of designing socially centric technologies is the explicit and implicit support of these protocols. A socially-centric platform therefore influences and supports the community of its users. For this reason, the design of socially-centric platforms is not socially (or politically) neutral. There is, for example, the question who is privileged (or discriminated) by the designed platform?

The process of agreeing objectives and choice of technology is cyclical or even continuous. Agreed common goals lead to the selection of technology, which, if implemented, modifies people's goals and expectations, which in turn leads back to the discussion on the selection of appropriate technology. Technology is becoming one of the social actors which influence or even take part in consulting and agreeing common goals.

[11] Negotiations understood as a resultant of many, sometimes contradictive, objectives and different forces and not as establishing a common position through discussion.

Among scientists active in computer science, the most promising approach is presented by constructivists. As opposed to substantivists, they do not assume that everything is already decided and, contrary to instrumentalists, they see the important and influential role of technology and information systems. The social constructivist theory sets, even if not directly, ambitious goals for computer scientists. Firstly, technology advancement broadens the set of possible solutions and applications, and secondly, even more importantly, pairing common social actors' goals, which are usually fuzzy and labile, with well-defined technological solution is a non-trivial task. Supporting users with a reliable and efficient technology which assures fulfilling predefined social goals in constantly changing environment requires:

- (semi)formal language to express and reconcile strategies, goals, contexts of social actors and social norms,
- widely accepted methods for measuring the level of satisfaction of the predefined social goals,
- resistance to smart, resourceful and adapting adversaries which are able to enter into coalitions and action groups,
- motivation mechanism encouraging actors to devote some energy to support common interests.

1.2.2. Social goals

The term *social goal* is used in two contexts: either as a description of personal objectives related to socialization and services to community (e.g. being a volunteer), or an aggregated and widely accepted joint objective of a community/society. In this dissertation the second meaning has been used. Social goals can be very different, depending on social system within which they exist, and can also change over time, but the most obvious are: justice, security, wellbeing or health. People are usually members of many different communities at the same time; some memberships are intentionally (e.g. author of this dissertation is a member of the ACM, the IEEE and the European Social Simulation Association because he filled out an application forms on purpose and paid membership fees); other communities are formed implicitly – e.g. people decide to join Allegro by creating there an account but do not have an awareness of being a member of community but rather they are focused on satisfying their own needs (e.g. using some services). In addition, some group of peoples (communities) with a common goal are created by chance – e.g. passengers travelling on a bus. Social goal should not be treated as identical with the reasons why people join communities.

The last case mentioned in the previous paragraph is very interesting because of a vivid example of contradictive objectives. The vast majority of travellers want to reach a destination in the shortest possible time but an adversary (i.e. pickpocket) has a different personal goal that is contradictory to the rest of passengers. Very often social goals are understood as maximizing shared values that are expressed either formally (by laws or rules) or informally (through culture). If we take a look at Wikipedia we can easily identify the most obvious goals shared by the majority of users. Objectivity, correctness and freshness are in the centre of all activities and discussions. Moreover, procedures and algorithms are made (or changed) to support the realisation of these objectives, e.g. articles tagged as controversial can only be changed by the privileged users to avoid vandalism and reverses.

Agreeing social goals of a community is not always an easy task and can take very different forms. Sometimes the process is carried out in the form of explicit negotiation between society members and the final result is reached by consensus. Sometimes, instead of by consensus, common objectives are set by voting. Some communities have fixed goals set by founders[12] and people either accept it by joining or reject by staying outside. Assuring social goals is usually limited in many different ways at the same time but the most commonly occurring dilemma can be often attributed to free-riders (or lack of cooperation). Using reputation systems will eventually bring some benefits (if the system is respectively well designed) but contributing to such systems requires time and energy and carries the risk of reciprocal negative evaluation. Moreover, research described in chapter 3.1. shows that more contributors make the reputation system more efficient. Therefore, the very common approach to developing socially aware algorithms is to encourage people to choose a cooperative strategy instead of selfish (e.g. P2P file exchange networks force people to not only download but also share some files what makes system more efficient and users' experience smoother by increasing the total bandwidth).

Regardless of the method applied to reach an agreement about expected social goals in a community, another problem seems to dominate the design of algorithms for realising social goals. Every methodological approach needs a way to measure the progress, thus every defined social goal has to be expressed in a computational form. Although it does not sound very complicated in practice it faces many problems. Attempts to develop a formal definition of equitable distribution are discussed in chapter 3.1. and show that even for a widely used and easily accepted terms measuring is not easy. More about measuring inequality, i.e. functions, measurements and discussion about questionnaires can be found in (Cowell 2011). The problem of measuring social goals is universal and present across many domains. Expressing in numbers on the community level such terms like health, wellbeing or security is non-trivial task. Good formal and widely accepted measures are lacking not even for social phenomena but also for security in information systems (Stoifo, Bellovin and Evans 2011). The most common solution is to combine many existing, yet flawed measures into complex multicriterial objective function. Such an approach is successfully applied in chapter 3.1.

[12] Setting social goals and trying to manipulate a society to achieve it is sometimes described as social engineering. This term has also been used by computer security specialists to identify an attack where psychological mechanisms are applied to convince people to break security policies (e.g. by revealing their own password by phone to an unauthorized person).

1.3. Data collection

1.3.1. Growing poll of data

The amount of data produced by each of us has grown exponentially in the last decade. Eric Schmidt, CEO of Google, claims that every two days we produce five exabytes of data, more than we did from the beginning of mankind to the early 2000s (Sigler 2010). Because the Google CEO did not reveal the methodology used for his estimation, some can question if it is realistic but a close look at Twitter shows that this most well-known microblog site generates every day between 15 and 20 TB of uncompressed data (Ryaboy 2011). Comparably, a large e-commerce site delivers on a daily basis over 50 TB of clickstream (Dietz 2009). Small computers, which probably by misunderstanding are still called mobile phones, accompany us in every step, in virtually every moment of our lives and register (or are able to register, which is an important difference discussed later on) sound, vision, location, movement, luminance and many more. Even quite a low-budget project can deliver hundreds of gigabytes data on a daily basis – a good example is "Human Speechome Project" conducted by Deb Roy and focused on registering the way infants acquire language (Roy 2006). The stream of data is produced by 11 video cameras and 14 microphones.

This dissertation is focused on social phenomena, therefore most examples are about collecting users' behavioural data. However, virtually every process can be a source of data – for instance, jet engine manufacturing generates ca. 3 TB data daily (Dietz 2009). The growing number of various sensors embedded into many devices along with the increasing precision[13] and ubiquity of the computers-intermediaries, which transfer all people's behaviours into electronic impulses, create a constantly developing and in the foreseeable future infinite pool of data. People now produce more data than we are able to either store or meaningfully process even though the storage costs decreased dramatically in the last thirty years (table 1.1) and a similar process applies to a computational power (table 1.2).

The dilemma stated in the previous paragraph forces website owners to choose a potentially valuable subset of data generated by users. This decision has to be made before we even know what is within the data, so it is usually based on either guessing or a try-and-fix approach. The infrastructure of the vast majority of e-commerce sites is unable to collect all data from clickstream even for a short period of time. Similar challenges are faced by astronomers who have already collected over 1 PB of data and it is estimated that this volume is growing at 0.5 PB per year (Berriman, Groom 2011). Thus, it is impossible to go back and look for some phenomena in the historical data as long as such data was not intentionally marked as valuable.

[13] The most obvious examples of growing sensor resolution are cameras embedded into mobile phones. The iPhone 4GS generates with every picture four times more data (8 million pixels) than the previous generation (2 million pixels).

Different strategies and heuristics can be applied to identify such subset of data and it will probably be a very interesting research area in the near future.

Table 1.1. Change of prices of one gigabyte storage in time (Doctorow 2011)

Year	1981	1987	1990	1994	1997	2000	2004	2010
Price of a Gigabyte (USD)	300,000.0	50,000.0	10,000.0	1000.0	100.0	10.0	1.0	0.1

Table 1.2. Growth of computational power in time (Wikipedia)

Processor	**kIPS/MIPS**	**Year**
Intel 4004 (740 kHz)	92 kIPS	1971
Intel 286 (12 MHz)	2.66 MIPS	1982
Intel 386DX (33 MHz)	11.4 MIPS	1985
Intel Pentium (100 MHz)	188 MIPS	1994
Intel Pentium III (600 MHz)	2,054 MIPS	1999
Pentium 4 Extreme Edition (3.2 GHz)	9,726 MIPS	2003
Intel Core 2 Extreme QX9770 (Quad core at 3.2 GHz)	59,455 MIPS	2008
Intel Core i7 Extreme Edition 980X (Six core at 3.33GHz)	147,600 MIPS	2010

Therefore, what data are in reality stored by companies? The answer received by Max Schrems from Facebook shed light on the question what companies know about us. The 24-year-old law student from Austria asked Facebook about all data collected about him during his three-year-long period of using the social network site. The reply he received is over 1200 pages of information, among them also some data that were supposed to be permanently removed and never answered invitations (AP 2011). It seems to be a large amount of data but compared to all possible data that can be collected during visiting a web page (among them all mouse-over events) it is rather small.

The amount of data and the need to select *a priori* most useful of them is only a beginning of challenges. Frequently, there is not a single point or a set of points managed by the same person or organization where all needed data can be gathered. The best examples are social network sites – every user can see only their own context within their own ego network. Therefore crawlers have to visit pages as different users and then a global view has to be reconstructed from hundreds of thousands of pictures. Such problem is valid not only for researchers but also for the biggest on-line players like Google.

Using many accounts of real users, which is the only viable option if a substantial subset of real data has to be collected, raises many privacy concerns and makes this approach a nightmare. Additionally, website owners extensively use legal actions against everyone (with the exception of Google and Bing) who tries to crawl their webpages.

1.3.2. Behavioural data

A constantly growing amount of data is not only a problem, but can also be seen as a great opportunity. For the first time in history, scientists have access to data about behaviours of entire communities. Moreover, data are collected exactly in the moment when events occur and are immediately accessible. Although every information system has a huge potential to gather data on users' behaviours, researchers see particular promise in mobile phones (or, to be more precise, smart phones). An increasing share of smartphones in the mobile phone market and the constantly growing computational power of ordinary mobile phones leads to the assumption that in a few years such a distinction, at least from data gathering perspective, will diminish). Raento even claims that mobile phones could be "the fMRI[14] of social science" (Raento, Eagle 2009).

Attempts to make sensing communities simple with the help of mobile phones (and accessible for people without programming knowledge) fail because of privacy concerns and the rapid development of mobile operating systems. One such framework is the ContextPhone[15] developed by the Context Group at the University of Helsinki. The pioneering works at MIT are still the reference point, especially that created by the way datasets are available to download from the webpage of the Reality Mining Project[16]. Recent smartphones allow you to collect much more data. Next to history of communication (i.e. calls and SMSs), and proximity-based Bluetooth, most Android-based smartphones can deliver some additional information listed in table 1.3.

Physical values are only a fraction of the possible data stream. A huge amount of information is generated by the interaction between the mobile phone and its user: a list of visited websites, gestures made, used applications, initiated/answered phonecalls or MMSs and SMSs. Interpretation of all that information and mapping in on appropriate behaviours and social situations is a non-trivial task. Changes in light intensity (sometimes bundled with proximity sensors used to switch off the screen during calls) can indicate when a mobile phone is in or out of the pocket. In the chapter 3.2. a framework is proposed to reproduce user's emotions based on data collected from a variety of sensors. Sensing mobile devices can be used not

---

[14] Functional Magnetic Resonance Imaging is used for brain imaging. Because this technology is affordable and non-invasive has a great influence on our understanding how human brain works.

[15] ContextPhone is an application working on Symbian 60 Series phones and can be used to collect information about users' actions (i.e. calls, SMSes, application in use, devices in range of Bluetooth).

[16] http://reality.media.mit.edu/ (03.2012).

only to conduct ethnographical and sociological research but also to manage natural disasters[17].

Table 1.3. List of additional information that can be collected form android-based smartphone

Information	Description
Location	There are many ways to obtain mobile phone locations, depending on the required precision, hardware and context. BTS triangulation works for all GSM phones, and GPS module delivers very good results when a phone can see satellites. With the help of databases a position can be pinpointed based on nearby WiFi access points. Additional techniques are now developed to support indoor localization (Merritt 2011)
Proximity	Bluetooth, WiFi and RFID can help to identify some close objects. One method to fix an exact position has been developed by Color Labs* – a start-up working on a technology which can tell if two people are close to each other with the help of pictures and background noise
Orientation in space	Gyroscope and electronic compass deliver information about device orientation and can also be used to trace movements (direction, acceleration, etc.)
Environmental variables	A range of environmental variables like temperature, magnetic field, light intensity or even humidity** are implemented by Android but usually not all are present in every particular device
Multimedia	Two cameras (front and back with high or even ultra-high*** resolution, or a sensitive microphone) can deliver a continuous data stream

* http://www.color.com/#landing (03.2012).
** Humidity sensor was added in the version 4.0.
*** A recent Nokia mobile phone presented during Mobile World Congress 2012 in Barcelona is equipped with a 41-Megapixel camera (Miller 2012).

Mobile phones are only one of many existing intermediaries used by people to manipulate the surrounding world, access information or maintain social contacts (some of them are even older than computers – e.g. credit cards). Nowadays, hundreds of millions of people spend long hours at the computer each day using web browsers. Not only a list of visited websites but also emails, chats, mouse movements, keystrokes and activities on social networking sites can be seen from the level of the web browser. Moreover, web browsers have access to information that is invisible for the Internet providers like the content of the websites transferred with the help of HTTPS protocol. On the other hand, the plethora of existing web technologies and websites makes access to some information from the web browser a nightmare.

[17] Innovative Support to Emergencies Diseases and Disasters (http://www.instedd.org) was founded in 2006 with a strong support from Google.org and the Rockefeller Foundation and develops open source solutions dedicated for developing countries and utilising the popularity of mobile phones to manage emergency situations and natural disasters (Ed 2009).

A referential system composed of a plugin for Chrome, and a server dedicated to collecting people's behaviours, was developed as a part of a bachelor project and supervised by the author of this dissertation. A detailed description of the developed solution can be found in (Czarnecki 2012) but some observations are worth mentioning here. External objects like flash animations or movie players either do not have an API that will allow reading their state from a plugin or, even if they do have such an API, it does not work properly. Therefore, the collection of data about videos watched by an Internet user on either YouTube[18] or TVP[19] required dedicated workarounds. Preparing a manually crafted solution, dedicated for a particular website, is not a viable option as long as we want to have 100% coverage. Nevertheless, focusing on the most popular sites will eventually deliver an acceptable performance.

Behavioural data are generally more appreciated than declarative data. It is quite common that users do not answer truthfully when filling questioners. Cheating by modifying behaviours is much more demanding for cognitive resources. On the other hand, the awareness of being observed can cause people to refrain from certain behaviours. Web usage patterns of workers who are aware that their employer monitors outgoing Internet traffic are certainly biased. Additionally, more and more people are convinced that they are constantly monitored on the Internet (which is to some extent true, especially when we take a look at solutions used for serving targeted ads) and thus they may either change their manner or do some techniques to obfuscate[20] it.

Although automatic collection of behavioural data eliminates some common problems with data quality, it should not be treated as an ultimate answer for all problems. Nowadays people often use more than one mobile phone or computer. We risk obtaining a blurred picture if we miss data from even a single intermediary. It is often difficult to guess whether the observed social phenomenon is an interesting result or the consequence of a lack of certain data. Another non-trivial task is the assignment of data collected from various sources to a specific user (especially when there is no strong authentication). Moreover, not all data, which at first sight appear as behavioural, are in fact behavioural. A list of friends in a social network site (e.g. Facebook) is a good example of declarative and not behavioural data – users explicitly click on a person and tag him/her as a friend[21].

In practice, there are two main sources of behavioural data. The dichotomy is based on either a centralized or decentralized approach. Huge, high quality, well-structured data can be obtained from service providers. Datasets containing

---

[18] http://www.youtube.com

[19] http://vod.tvp.pl

[20] Tools and techniques used to hide oneself in the Internet are beyond the scope of this dissertation but let me name a few: "do-not-trace" button and "privacy mode" in web browsers, proxies, onion routing (e.g. Tor – https://www.torproject.org/).

[21] Behavioural data about friends can be collected from the history of communication (i.e. chats, emails). Research in which the author of this dissertation was involved shows that behavioural data from an instant messenger can be used to predict the type of relationships between interacting people (Doniec, Hupa and Nielek 2009).

behaviours of millions of users (players, sellers, buyers, web 2.0 content producers, etc.) are not rare. This kind of dataset is extensively used in this dissertation to test some reputation algorithms supporting fairness (see chapter 3.1.) and to mine for interesting social phenomena (see chapter 2.1.). Although datasets from service providers are invaluable for research, some limitations also exist. Firstly, because data about user behaviours are crucial for companies (and even more for their competitors) it is really difficult to obtain such a dataset. Secondly, production systems are usually tuned to collect only the most important data for delivering smooth and high-quality services, therefore some interesting information from a scientific point of view may be missing. Thirdly, the dataset is limited to one type of activity (e.g. we know how a person behaves as a player but we cannot match this information to their buying history on eBay). On the other hand, collecting data in a distributed way, close to users, allows taking a look across different activity areas and obtaining almost all imaginable information. The main problem with the second approach is that it is virtually impossible to convince millions of people to take part in our study. Researchers offer some incentives[22] to woo people but it carries the risk that those who take part in the survey will not be a random sample of the population.

1.3.3. Text mining

Scientists use term *unstructured data* to describe information expressed in natural language without any additional tagging. Blogs, posts, textual reviews, comments, articles, videos and most websites are examples of *unstructured data*. The most commonly used way to express opinion and provide information is also very inconvenient for automatic processing. As long as computers do not understand[23] the meaning of information services offered are very limited and often have disappointing quality. Answer the question "In which city is the highest building in Poland?" requires either a database of buildings in Poland with all features like height and location or a way to understand searchable information in Internet.

Creating and updating databases containing large number of objects is a costly and tedious task[24]. On the other hand natural language processing (NLP) is an AI-complete[25] task. Therefore, the third approach that combines the previous two has

[22] Virtually everything can work as an incentive but it is becoming increasingly popular to offer interesting and useful services (e.g. reputation evaluation, recommendation etc.).

[23] The term *understand* in a context of computer often raises a fierce debate involving philosophers, cognitive psychologists and AI specialists. Leaving aside question of whether understanding is an integral part of intelligence let assume for the need of this dissertation that the term *understanding* means *"being able to combine and manipulate information with result which is meaningful for people"*.

[24] However, some companies decide to follow this path with a good results (e.g. Wolfram|Alpha – http://www.wolframalpha.com/).

[25] The term AI-complete, firstly used by Fanya Montalvo, describes the difficulty attributed to natural language processing. According to Regina Brazilay *"all difficult problems in artificial intelligence manifest themselves in NLP problems"* (Barzilay 2010).

been proposed. Resource Description Framework[26] (RDF) proposed by World Wide Web Consortium (W3C) was designed to deliver a common standard for semi-automatic tagging of the web. The idea of semantic web[27] used to be considered as the next, natural step in the evolution of the Internet. However some proponents of this technology are very optimistic, and despite the support of recognized authorities and over twelve years that passed from the first publication of the RDF standard, proposed solution gains ground very slow and with great difficulty. Therefore, we still do not see (and will not see in a foreseeable future) comments on the Facebook or Twitter, which are RDF compliant. Wilks noticed this problem and proposes to use NLP algorithms as a basis for developing semantic web (Wilks 2010).

In parallel, many works were (and are) pending on the efficient and precise understanding and processing algorithms for textual information expressed in natural languages. The top-down approach in which linguists have tried to build a formal model of language has proved to be too complex and practically inapplicable. Nevertheless, attempts were also made for Polish (Przepiorkowski et al. 2002). On the other hand a successful application of machine learning algorithms to solve a variety of problems brought hope to the good results in natural language processing tasks. From over ten years the vast and growing number of published papers shows the use of machine learning algorithms to text classification (Gantner and Schmidt-Thieme 2009), summarizing (Aone, Okurowski and Gorlinsky 1998), automatic translation (Menezes 2002) and many more. Nowadays, natural language processing seems to be inextricably linked to machine learning algorithms.

In general, the use of domain knowledge improves the precision of machine learning algorithms. That is also truth for natural language processing tasks. A hybrid approach usually assumes that a language specific knowledge is used transform text into semi-structured form, which is then processed by machine learning algorithms. Transformation can be rather simple and based only on a list of substitutions and a limited dictionary (as is shown in chapter 3.2.2. by analysing SMSes dataset) or very sophisticated with shallow parsing tools, diactric guessers and many more (as is shown in chapter 2.1.5. by mining the meaning of comments). The main problem with an approach combining machine learning with the knowledge about language structure and properties is the need for crafting a separate solution for each language independently. Grammars patterns, dictionaries and term frequency information acquired for English cannot be used for Polish.

Although, up-do-date dedicated NLP algorithms deliver a really good performance (see chapter 2.1.5. for some benchmarks) all users should be aware that it is a statistical tool. Most applications, with an automatic translation working as an exception, assume a decent size of dataset and return results, which can be safety interpreted only on the aggregated level. Nonetheless, natural languages processing

[26] http://www.w3.org/RDF.

[27] Tim Berners-Lee, inventor of World Wide Web and President of W3C, coined the term *semantic web*.

tools already left research labs and find successful commercial applications and are used for sensing social phenomena.

NLP tools are crucial for identifying and discovering regularities and even social phenomena in textual data accessible in information systems. Latent Semantic Analysis was applied to over one thousand SAVE Award[28] submissions and SMSes sent to the President of the United States and is considered as a vital tool for developing e-democracy (Evaangelopoulos, Visinescu 2012). More business-oriented approach is presented by Bluefin Labs[29], a start-up founded by Bed Roy, former director of the MIT Media Lab's Cognitive Machine Group. A stream of social media content (mostly Twitter and Facebook) is collected in an almost real time, processed and compared with broadcasted TV shows and TV ads. Relevant information are extracted from textual information and matched with a specific statement/event on TV show or ad (Talbot 2011). For the first time technology allows sensing the opinion of spectators in real time and in a scale inaccessible up to now.

The amount of accessible data and the fact that more and more human activities are reflected in the Internet make sensing complex phenomena engaging the whole society possible. An attempt to predict prices of shares on the Warsaw stock exchange that based on the data crawled from the Internet on a daily basis is presented in chapter 2.2. Similar works with not so distant results have also been done for the New York Stock Exchange (Tetlock 2007). Even if processes affecting the shares' price can be seen as elite[30] and hermetic to some extend the same complaint cannot be raised with regard to political view and the support of the political parties. Research published by the author of this dissertation shows that it is possible to predict the outcome of the election basing only on the NLP tools and information crawled from the biggest polish news sites (Wawer, Nielek 2008).

1.4. Social simulation

1.4.1. Introduction

The dynamic development of the simulation started with the advent of increasingly powerful computers in the early 1980s. The possibility to build a virtual model of almost everything became available to a vast majority of researchers and practitioners. Next to an early implementation in aeronautics and defence, appeared many models in physics, car design and construction. Simulation is considered to be one of the

[28] Securing Americans' Value and Efficiency Award (SAVE Award – http://www.whitehouse.gov/save-award) was established in 2009 by Barack Obama to encourage ideas making U.S. government spending more efficient. The award was enthusiastically received by civil servants and resulted in more than 18,000 submissions.

[29] http://www.bluefinlabs.com.

[30] Some researchers argue that stock exchange indexes are a good measure of a current mood prevailing in the society (Casti 2010).

most influential forces (next to new materials) in shaping contemporary architecture – computer-aided design has enabled the design of structures with innovative and unusual shapes. The majority were content with the new and powerful tool but some scientists have raised concerns. Physicists from the most famous universities, among them MIT, professed many arguments against simulation, starting from the philosophical problem of absolute truth and finishing with concerns about the influence on the educational process (Turkle 2009).

A broad use of computer simulation to study social phenomena appeared much latercompared to the general application. Very early works pointing to the interesting intersection between social science and simulation were published in the early and mid-90s (Gilbert 1991; Troitzsch 1995) but a more systematic look was presented by Nigel Gilbert and Klaus G. Troitzsch in the book "Simulation for the Social Scientists"[31] a few years later (Gilbert, Troitzsch 1999). Around the same time the Journal of Artificial Societies and Social Simulation[32] – the most prominent forum for exchanging ideas and presenting results of social simulation – was established and was followed by forming the European Social Simulation Association[33].

It is worth mentioning that very early attempts to build formal models of social phenomena were taken even before the first computer was built. In 1909 Agner Krarup Erlang published a paper proposing queuing theory and almost twenty years later John von Neumann presented game theory – an elegant framework for modelling individuals' interaction. Two centuries earlier Daniel Bernoulli[34] used differential equations to describe the process of spreading disease in population (Dietz, Heesterbeek 2002).

Nevertheless, computer-aided social simulation is a relatively young discipline, so researchers devote a lot of time to epistemological issues and questions. The issues mentioned vary from a very fundamental one like "in which interrelation are simulation and observable data" or "what emergence really means" to more technical ones e.g. "How to agree a terminology" or "the consequences of finite precision". A good review of different positions and problems can be found in (Epistemological Aspects of Computer Simulation in the Social Sciences 2006). Even though social simulations use mainly the same tools as simulation of physical phenomena, they vary greatly in the interpretation of results and the place they occupy in scientific reasoning.

---

[31] This book was also the first and for quite a long time the only textbook for social simulation courses. Only recently a publication of "Agent-Based and Individual-Based Modeling" by Steven F. Railback and Volker Grimm has changed this situation.

[32] http://jasss.soc.surrey.ac.uk/JASSS.html.

[33] ESSA organizes an annual conference gathering the most active researchers in social simulation (http://essa.eu.org).

[34] The Bernoulli family has their own place in the history of mathematic. Daniel Bernoulli was a nephew of Jakob Bernoulli – creator of probabilistic theory – and son of Johann Bernoulli – educator of Leonhard Euler.

A unique characteristic of social simulation is their interdisciplinary approach. It functions as a meeting point for researchers from many disciplines and each group bring their own tools and methods. Physicists apply differential equations and complex models of physical phenomena to explain people's interaction and behaviours. They are able to find similarities, even if distant, between opinion formation and mathematical model of ferromagnetism (Stauffer 2002). A good example is the Sznajd model, which is based on Ising spins model[35] and was proposed as an explanation of the level of political support for particular candidates in election (Sznajd-Weron 2005). Sociophysics and econophysics are more and more popular but are also subjected to a fierce critique. A brilliant paper, which pinpoints gaps in connections between social science and physic, carries a self-descriptive title: "Call for Closer Link with Reality" (Sobkowicz 2009). At the same time, ecologists who have a long tradition of applying differential equations (e.g. epidemic or predator-prey model), abandon them in favour of individual-based modelling because *"instead of thinking about population that have birth and death rates that depend only on populatioemogn size, with IBE (individual-based ecology) we think of systems of individuals whose growth, reproduction, and death is the outcome of adaptive behaviour."* (Grimm 2005).

Agent-based modelling and differential equations do not cover the entire set of available tools. For example, very interesting studies of linguists are conducted with the use of cellular automata (Beltran et al. 2009). Microsimulation is often used for traffic modelling (Decoster 2011) or tax income prediction (Claes 2009). Traffic management and taxation policies are in fact one of the areas where social simulation results are directly applicable in practice. Apart from them social simulation results are often applied to solve practical problems in disciplines like modelling and prediction of water or power usage, urban development, demographic modelling and epidemiology.

1.4.2. Relation between social simulation and real world

Models are built to resemble real life and to yield some knowledge about it. Such a simple claim fires a longstanding argument between those who think that modelling is a very convenient tool or even a *"third way of doing science"* (Axelrod 1998) and their opponents who are, like Cartwright, of the opinion that model is *"a work of fiction"* (Cartwright 1999). Neither of these strong positions is entirely justified. Oreskes in her well-received (but also controversial) paper published in Science (Oreskes 1994) gives a strong support to the thesis of very limited usability of numerical models and simulation. Although the arguments recalled by Oreskes are biased in one direction and are mostly based on her previous experience in earth science, they are also convenient for settling discussion. The main accusations recalled in the paper are summarized below:

[35] Ising model was invented by Wilhelm Lenz and defines a mathematical model of ferromagnetism.

– closed system does not exist in real world and, thus we are unable to verify a model because *per definition* it is closed,
– laws in general, and those used in building model in particular, are only a simplified version of reality; thus we miss some details which may be crucial for obtained results,
– parameters cannot be measured with infinite precision, thus at the beginning we must accept some deviation from the real world,
– resolution of a model usually does not fit to data resolution thus we either have to aggregate or interpolate some points to match available data with model resolution.

Despite these reservations, Oreskes sees some applications of models and proposes to use modelling for challenging existing formulations. Although the difference between being able to confirm something and being able to rebut is obvious and often has very far-reaching consequences, in many practical cases where the task is to choose one of several competing theories explaining particular social phenomenon both approaches can be successfully applied and lead to similar results.

It is worth noticing that many strictures of the opponents of simulations are *de facto* rooted in the chaos theory and can be easily addressed by the analysis of a system stability. Since the early 60s and the pioneering works of Edward Lorenz it has been known that for certain classes systems prediction, and thus modelling and simulation, is impossible because of its huge sensitivity to initial parameters. A broad selection of real systems behave quite stably, at least for a subset of possible initial parameters, thus a useful and practically applicable model has been built. Computer models of new buildings, planes or cars have proved to be efficient in limiting, but not eliminating[36], the need for building and testing real objects.

Both Oreskes and Cartwright expect the truth from simulation and modelling, but the goal of scientific theories is not always the truth but empirical adequacy (van Fraassen 1980). Isaac Newton' gravitation theory is a very good example of a scientific theory which, as we already know, does not describe the real nature of gravitation but because it delivers a very precise results for average size and speed, it is still very often used instead of Albert Einstein's general theory of relativity. Hence, from the practical point of view, absolute truth should not be treated as an exclusive objective of science. Fraassen (van Fraassen 1980) points out two clear epistemological positions:

– scientific realism – *science tries to deliver theories which are true; truth is the only factor used to either accept or dismiss theory,*
– constructive empiricism – *science tries to deliver theories which can be verified against empirical data; empirical adequacy is the only factor indicating whether to accept this theory or not.*

[36] A development of computer simulation techniques causes the practical elimination of the nuclear weapon test explosions. Since early 90's Russia, USA, and the United Kingdom have not conduced any nuclear test explosion. France and Chine conducted their last tests only few years later.

The question whether the two approaches listed above rule out each other remains open. Firstly, a matching between real data and theory increases the chance that theory pinpointed a phenomenon, and secondly, a theory which is true should (with some rare exceptions[37]) deliver a very good matching between observable data and theory. On the other hand, an Austrian-born philosopher Karl Popper and the author of the term "critical rationalism" stresses the importance of falsification in favour of validation. He claims that even a big number of positive outcomes cannot confirm a scientific theory, but even a single counterexample can falsify it.

Almost all disputes regarding the applicability and usability of the simulation relate also indirectly to the social simulation but non-mathematical formulation of the vast majority of social phenomena raises many additional concerns. Differences between social simulation and physical simulation are one of crucial factors fuelling epistemological discussions. Following a list of distinctive features proposed by Rossiter in (Rossiter 2010) four substantial differences can be identified:

– **Quantitative vs. qualitative matching to reality**

Practically all but small subset of social phenomena models devoted to epidemiology are constructed to fit real data only on a qualitative level. Observable phenomena in models and states in system match the ones existing in real life but it is usually difficult to exactly identify changing points.

– **Strong vs. limited predictive accuracy**

Opposed to simulation of physical phenomena, where a prediction precision is limited mostly by computational power[38], in social simulation two additional factors limit the horizon of simulation – lack of quantitative description of people's behaviours and their stochastic nature. Naturally, both types of simulation suffer from deterministic chaos but it is still easier to measure temperature or distance than the moods of people.

– **Universal mathematical laws vs. context-specific social factors**

Simulation of physical objects and phenomena is based on well-defined, computable mathematical laws. In social systems such well-defined laws either do not exist or we do not know them. Thus, every model is by definition dependent on many contextual information and social factors.

[37] Sometimes an evaluation of the matching between real world and theory can be difficult. Good example is a situation when more than one theory deliver a good, and empirically comparable adequacy. Another example is when the proposed theory is too complicated to compute.

[38] This statement is true as long as objects of "reasonable" size and weight are considered. On the level of atoms, where quantum mechanics and statistical laws play the first role, physical phenomena are very unpredictable. It is impossible to answer which atoms will decay in the next hour but at the same time it is pretty easy to guess the level of radioactivity of a sample of such atoms even in a very long term. This possibility is used, among others, to determine the age of object with help of radiocarbon dating.

– **Occan's razor vs. "inherent" complexity**

When simulating results of a car crash test or the efficiency of a new plane it is quite simple to draw the line between relevant and irrelevant factors. The same line in social simulation is blurred or non-existent. Therefore, the need to take into account many factors makes the design and understanding of social simulation much more complex.

Some of the limitations mentioned above can be minimalized (or even eliminated) by the use of the growing pool of behavioural data. The opportunity of constantly tracing people's behaviour can lead to more complex models and evento the qualitative instead of quantitative matching to reality.

Differences not only substantially increase the complexity of social simulation but also change the meaning of terms like: validation, verification and calibration. Instead of trying to prove that the model is true, the term *verification* in social simulation is used to confirm that implementation is correct to researchers' intentions (Edmonds 2003). Such definition requires a comprehensive and semi-formal approach for model verification. Firstly, formalism is needed that allows a computable expression of expectations and intentions of the simulator. Secondly, a rigorous approach to testing and debugging process is required, and thirdly, common used frameworks like Repast Simphony[39] or MASON[40] have to be established and be free from major bugs. But even a bug-free implementation does not guarantee a smooth simulation. Some practitioners emphasize that many models are susceptible to floating-point arithmetic and so-called "ghosts"[41] (Polhill, Izquierdo and Gotts 2005; Izquierdo, Polhill 2006).

Contrary to verification, the process of validation relates to the relation between model and social phenomenon. Validation should either confirm or deny that the model resemble real world. Naturally, the degree of similarity and features which match the model and social phenomena depend strongly on initial assumptions. Calibration, the third term used next to verification and validation, is more a part of model developing process than an evaluation process. An interesting approach to calibration is presented in (Grimm 2005) and a good example of validation of demographic simulation can be found in (Montanola-Sales, Onggo, Casanovas-Garcia 2011). On the margin of calibration it is important to remember that building a very complicated model with a lot of initial parameters and, at the same time, having only a limited amount of data from real world to compare may cause the situation where model can be fitted to virtually every set of data and every social phenomenon. Only

[39] http://repast.sourceforge.net.

[40] http://cs.gmu.edu/~eclab/projects/mason.

[41] The term "ghosts" is used in simulation to describe agents which, usually because of errors in rounding and the nature of floating-point arithmetic, are beyond intentionally projected states for agents. For example, it can happen that a fairly simple operation will return a very small number instead of zero that was expected and thus agent will be neither dead nor live.

keeping the model relatively simple makes the results of validation and calibration plausible.

All the limitations mentioned above do not disqualify simulation as a tool used in the study of social phenomena but require a careful understanding of areas of applicability and an ability to define reachable goals and viable expectations. Social simulation can be used especially when:

- we know the final state and we want to identify forces which shaped it,
- we know existing processes; we can observe and describe them but we want to know the final, long-term outcome,
- we want to test a new solution which theoretically can be done experimentally but because of costs such approach is inappropriate,
- we cannot engender on purpose some situations because they are unethical or dangerous for people.

A well-defined problem, lack of exaggerated expectations and a very focused research area are a precondition of a successful application of simulation to social science. Those requirements are independent of the techniques used. Regardless of whether you apply differential equations or cellular automata or agent-based modelling, the main limitation and difference remain the same but, of course, can be complemented by specific features of the selected tool. Cellular automata do not allow differentiating actors (everyone has the same properties), practical implementation of differential equations is limited to the pretty simple models and agent-based models, which are the most universal, require an extensive initial dataset, which makes mapping real life to model viable.

Successful applications of social simulation combine collected datasets about observable behaviours and knowledge from many disciplines. Agent-based social simulation[42], which is the most popular approach among scientists, and is also extensively used in this thesis (see chapter 3.1.), is an intersection of different fields ranging from computer science to psychology and behavioural biology. Not only does a model have to be formally described and implemented but it also has to be plausible in a matter of social interaction, group dynamics, etc. Some researchers see the agent-based approach as a bridge between researchers with different backgrounds and a tool *"for providing cross-fertilisation"* (Davidsson 2001). While the agent-based approach helps to organize discussion about interactions between individuals, other tools are needed to facilitate collaboration in modelling and formalizing of intra-agent processes (e.g. heuristics used by agent to take decisions).

[42] Biologists and ecologists use the term "individual-based" instead of "agent-based", which stresses the importance of single individual and specifies the granularity of simulation.

1.4.3. Emergence

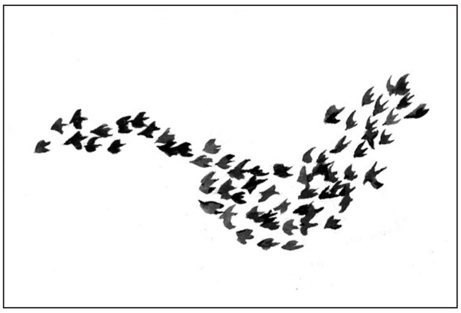
Very early deliberations about the term *emergence* can be attributed to Aristotle and followed by John Stuart Mill, Julian Huxley or G.H. Lewes. Leaving aside for a moment the formal definition and philosophical debates, emergence can be described as a phenomenon in which a sum of individual, rather uncomplicated behaviours leads to complex phenomena on the higher level. The properties and the nature of those complex phenomena cannot be easily inferred from a simple interaction. Figure 1.2 shows a few pictorial examples of complex phenomena that emerge from a simple behaviour. The common feature of all presented phenomena is the lack of either central guidance or direct communication between individuals.

Emergence is not just birds, cars and termites (or ants). Another good and often cited example is the relation between neurons and brain. Complex brain functions like imagination or intelligence are not a trivial aggregation of neurons' activation function and their connections. Many researchers, among them Robert Axelrod (Axelrod 2006), also agree that evolution of cooperation is an outcome of simple strategies on the individual level and not a result of top-down approach.

Many philosophical theories have been created to explain the nature of the additional factor, which sometimes appears during the addition of simple, individual behaviours. Some philosophers even claim that nothing like a special effect exists and the whole is only a sum of parts and we cannot decompose complex macro effects only because of our lack of indispensable resources and knowledge. Searle identifies five types of reduction (Searle 2008):

- ontological reduction – *most general form of reduction; it is based on the assumption that an object of certain type consists of nothing more but a collection of other objects,*
- property ontological reduction – *similar to ontological reduction but focused on properties of objects, e.g. light, smell, heat,*
- theoretical reduction – *ability to show that one theory can be explained and reduced to another theory, e.g. Newton' gravitation theory can be reduced to Einstein' general theory of relativity,*
- logical or definitional reduction – *relation between sentences (words), where one sentence can be translated into another without any losses of meaning or precision,*
- causal reduction – *it is similar to ontological reduction but focused on casual consequences of objects; casual property of an object can be explained by casual property of reducing phenomena.*

Neither philosophical discussion nor the dispute between reductionists and supporters of emergency are the main scope of this dissertation. Those who want a closer look at the history of these disputes can look into "Social Emergence" written by R. Keith Sawyer (Sawyer 2005). For scientists active in the area of social simulation the most important questions are not those about the nature of emergence or their epistemological background but whether we are able to "grow" a complex system from a quite simple heuristics attributed to individuals.



a) bird flock[43]

b) trafic jam[44],

c) crowd sitting in hall[45]

d) termite mound[46]

Fig. 1.2. Examples of emergent phenomena in real life

Examples of generative approach to science, rooted in Leibnitz monadistic philosophy, can be found in the early works of American linguist Noam Chomsky and his generative grammar but the most influential motto, from social simulators' perspective, was stated by Joshua M. Epstein: *"if you didn't grow it, you didn't explain it."* (Epstein 1999). For generativists the indispensable condition for understanding the process is the ability to replicate it and build it from small pieces of individual behaviours – virtually from scratch. On the other hand, it is important to avoid a quite common misinterpretation of the Epstein motto. The ability to grow a complex

[43] Each bird follows its immediate predecessor, which is seen by him. Such a simple rule enriched with a few parameters is enough to create a complex bird flock.

[44] Thomas C. Shelling in (Schelling 1978) describes the situation where an accident in outbound lane jam up the city bound traffic. Every driver passing the wreckage on the opposite site of the road slow down to take a look and, thus, limits the bandwidth of the road. Drivers collectively "decided" to spend many extra minutes waiting in exchange for a few seconds look.

[45] This example is also taken from Thomas C. Shelling (Schelling 1978). He examines, inspired by his own experience as lecturer, the way people take places in the hall during lectures. Simple rule, which can be in short summarized as the need to socialize, follows to strange and non-optimal results.

[46] Mitchel Resnick in his textbook (Resnick, Zeckhauser 2002) shows how to grow up a termite mound with help of the NetLogo. Termite mounds are another example of a complex structure, which emerge from a simple action of blind, small insects without a central coordination.

social phenomenon from simple individuals' strategies does not necessarily explain why people follow particular strategies, what motivates them and what assures that selected strategies are evolutionarily stable.

Evolutionary biologists and sociologists very often use the term *emergence* in a slightly different meaning described in the previous paragraphs. A comprehensive answer to questions like how language emerged or why we observe altruistic behaviours in society requires not only the identification of individual strategies but plausible proofs that such strategies could evolve and their existence is justified from individuals' perspective. Emergence is often interrelated in social context with evolution.

1.4.4. Practical realisation

In the next chapters social simulation is used to verify the proposed algorithms. Many details about practical implementation in a particular context and steps, which need to be taken into account to obtain credible and easy to interpret results, are at length described in chapter 3.1. but some observations and strategies are more universal and can be applied across many subjects and are thus collected in this section.

A comprehensive checklist for conceptual design of multi-agent system can be found in a book written by two ecologists Volker Grimm and Steven S. Railback (Grimm 2005). The authors identify ten main points in developing social simulations:

– *Emergence* – virtually all interesting social simulations contain some emergent properties; even if the main goal of simulation is not focused on these properties it is crucial to know where complex phenomena on the macro level can be expected (it is often used as an indicator of correctness of implementation),
– *Adaptation* – do agents change their behaviour in time and if yes, what is the mechanism which forces them to do so?
– *Fitness* – for all simulations resembling evolution definition of a fitness function is crucial; evolutionary pressure can be modelled in many ways but the most commonly used approach is based on some kind of energy,
– *Prediction* – expected precision, granularity and time horizon have to be fixed,
– *Interaction* – in a vast majority of simulations there is a kind of interaction between agents (either direct or indirect with changes of environment as an intermediary),
– *Sensing* – agents have to have some knowledge about other agents and environment; How do agents acquire this knowledge? Is this mechanism error-prone and how costly is it for individuals?
– *Stochasticity* – all stochastic processes have to be identified and their influence on the level of stochastity of the model should be estimated (it is required to plan a sufficient number of repetitions of experiments),

- *Collectives* – do some collectives exist in simulation? How are they created? Explicit as a separate object with some properties or implicit as a sum of smaller, independent objects?
- *Scheduling* – computers cannot natively use a continuous time, so all computer simulations are based either on discrete time or discrete events; next to time representation and their granularity in simulation a researcher has to plan the order of execution and mimic simultaneity of real events,
- *Observation* – every simulation should have an observation plan; with regard to what effect we want to research some values have to or do not have to be saved for further investigation.

Although the checklist proposed by Grimm and Railback can be used across many disciplines, some bias in direction of ecology is noticeable. They omit some fundamental questions for everyone who wants to do design a social simulation, like space or granularity. For biological models properties of the real system usually determine decisions about shape and dimension of the designed model. Scientists designing models of social phenomena tend to have more freedom in this decision. Modelling social interaction with the help of social network can eliminate the need for representation of physical space in the designed model. Sometimes a complex three-dimensional space can be reduced to only two dimensions, which are much easier to model and analyse (e.g. modelling a lift in a building does not require a reflection of the whole building).

Researchers have yet another problem with choosing the appropriate scale of a model. The obvious solution is to use one-to-one matching between the real system and model but in virtually all cases such an approach is either impossible or inappropriate. None of the researchers have enough computational resources to develop a simulation of spreading of a disease in which every single human/animal in real world will have a counterpart in the model. Two approaches to work around this problem are most common. Modellers can either scale down their models by choosing one of ten/hundred/thousand individuals (complex and non-trivial is the question how to make such sampling) or group some individuals and represent them in simulation as a single agent. Simulations of buyers and sellers in auction house presented in chapter 3 are done in a scale of one to two hundred (because most users usually appear only in one function – either buyer or seller – only small differences in scaling exist for both groups of users).

A limited size of the model will help by keeping the whole simulation simple enough to manage the interpretation of results but alone is not sufficient. Natural willingness to add new properties and factors during the designing process has to be limited in favour of deeper understanding of smaller subsets of phenomena and processes. The fact that we are able to code particular heuristic does not necessarily means that we will be able to interpret a complex outcome of agents' interactions. Therefore, all simulations presented in this dissertation are relatively simple on the conceptual level. Instead of multiplying heuristics and individuals' properties, which will differentiate agents, the greatest emphasis was placed on the theoretical foundation and result interpretation. Simple models require also less time and effort from research community to gather it, thus are easier to replicate, widely accepted and more often cited. Simplicity is usually a strong prerequisite for correct simulation.

Making a simulation model too complicated opens yet another issue. A growing number of initial parameters requires more observation data to conduct tuning. Moreover, if the number of parameters is high, compared to a relatively simple function used for tuning, fitting between real system and model can be reached by chance instead of by good resemblance to reality. An increasing number of function parameters leads to the situation where such functions can mirror virtually any social process. Additionally, cartesian products of parameters make sensitivity analysis of social simulation with even an average size of parameter space computationally difficult.

Although almost every programming language can be used to develop models and conduct simulation, a vast majority of practical implementation uses only a few frameworks. The most advanced and, at the same time, the most popular are MASON[47] and Repast Simphony[48] (both developed in Java and, thus, platform-independent). Next to these two frameworks at least one more is worth mentioning – NetLogo[49] is often used for teaching and also fast prototyping. Dedicated solutions provide very convenient features and configurable elements, which greatly accelerate the whole process of designing and implementation but some serious social simulations were also developed using Excel. Spanish scientists used cellular automata implemented in Excel to forecast language shift in one of Spanish provinces (Beltran 2009). Those who need a simultaneous simulation of many hundreds of thousands of agents should also take a look at Erlang[50].

Existing frameworks greatly decrease the amount of work and required programming skills for developing models but because social simulation is an interdisciplinary approach and thus is commonly used by researchers with diverse backgrounds

[47] MASON stands for Multi-Agent Simulator Of Neighbourhoods and is a framework for discrete-event multiagent simulation written in Java at Georg Mason University. MASON is distributed as an open source and can be downloaded from: http://cs.gmu.edu/~eclab/projects/mason.

[48] Repast Simphony is also written in Java and offers similar functions to MASON. Repast Simphony allows you to create simulation not only written in Java but also in Groovy or NetLogo. A special version of the framework exists to perform simulation on clusters and supercomputers. Repast Simphony and Repast HPC are published on "New BSD" licence and can be downloaded from: http://repast.sourceforge.net/index.html

[49] NetLogo is described by its authors as „a multi-agent programmable modelling environment". It is a user-friendly, simple, yet powerful framework for developing multi-agent simulations by a mix of drag-and-drop approach and programming in a thoroughly modified version of the Logo language. NetLogo is written in JAVA and, thus, is platform-independent. Source code and binary version can be downloaded from: http://ccl.northwestern.edu/netlogo.

[50] Erlang is a functional language developed in mid-80s by Joe Armstrong during his work in Ericsson laboratories. From early beginning Erlang was designed to support massive multithreading (Erlang's virtual machine handles easily many hundreds of thousands of threads simultaneously). Erlang supports only messages passing as a mechanism for communicating between threads (no shared variables), thus it is easy to map a conceptual interaction between agents on the designing specification. From 1998 Erlang is available on the open licence: http://www.erlang.org

(some of them cannot program at all) it still creates some barriers. Therefore a trend to develop purely visual tools for building models can be observed (e.g. NetLogo is equipped with Behaviour Composer[51] and Repast Simphony contains Repast Flowchart module). Visual builders usually restrict users to develop a model on the basis of one of predefined, yet flexible, patterns. It may (or may not) influence results but, simultaneously, makes analyses and model dissemination easier.

Regardless of the approach, whether we use a dedicated framework or visual builder or general-purpose language, any serious analyses of social simulation results have to be predated by a thorough check of the correctness of implementation. New, surprising results can be a great achievement but also a mistake in both implementation and data collection. The typical procedure of debugging social simulation code is quite similar to debugging any piece of software but some aspects require a different, more flexible approach. The main problem is that the outcome of the social simulation is in general not known *a priori*, thus comparing the simulation outcome with an expected behaviour requires some additional assumptions. One of the possible solutions is to identify extreme cases in developing model (e.g. in the reputation management system it is a situation where no one sends any reports about users' behaviours). Extreme cases generate results which are easy to predict and, thus, can be used for testing the correctness of implementation.

Social simulations done for research purposes are usually developed incrementally. Assumptions which are initially very simple, are extended with new concepts and features. Such approach makes it possible to thoroughly test the implementation at the early stages when the system is not yet so complex. Incremental development is often connected with a lack of planning and non-existence of documentation. Both facts significantly influence the quality of code and make the final solution sometimes virtually unbearable. A good habit is to split the developing process into two phases – first when many hypotheses are verified and software is implemented with fewer rigours and second when most promising hypotheses are already selected and software should be developed in an organized way and usually from scratch.

A broad class of social simulation systems is prone to errors due to float point arithmetic. Multiplication and subtraction can lead to surprising results, i.e. subtraction of two "should-be-identical" values can return a value close to (but not) zero. Thus, some agents can be in a state which is beyond what can be sensible interpreted. The ultimate solution is to abandon the floating-point arithmetic to a fixed point. Although many models can be implemented without float-point arithmetic, it is more complicated to implement and therefore is not a viable option[52].

---

[51] The official project web page can be found here: http://blogs.oucs.ox.ac.uk/modelling4all/. A very interesting video showing advantages of visual modelling can be seen on YouTube: http://www.youtube.com/watch?v=8AAZeTOj7yc&feature=youtu.be

[52] Sometimes researchers abandon the floating-point arithmetic during implementation for performance reasons. Parallel computation on GPU (e.g. nVidia CUDA) is most efficient with the fixed-precision arithmetic.

Another solution, which can be implemented and does not require too much additional work, is to carefully check agent's state every time it changes. Such a strong control in connection with a conversion to integer (every time when needed) reduces the risk of floating-point ghosts in simulation.

Even a formal proof of correctness of the implementation will not make simulation convincing to researchers who, like Oreskes, see the main problem in limited, closed nature of such models. A vast majority of models contain some simplifications[53]. In particular, models of social phenomena, because of their rich social context, are usually prone to limit the external interconnection. The real world is infinite as opposed to simulation models, which are always finite. Although the contradiction pointed out in the previous sentence cannot be eliminated, some methods exist to limit its consequences for the quality of research. During the designing stage all strongly connected elements should be identified and included into the model. On the other hand, all elements, which are only loosely connected with the most crucial part of modelled phenomenon, can (or even should) be omitted. In other words, researchers struggle to maximize the ratio between internal and external interdependencies. This approach can be applied universally in all designed simulations.

Sometimes, when behavioural data are available, researchers extend the approach described in the previous paragraph by using trace-driven simulation. The basic idea behind trace-driven simulation is to exchange an external link with previously collected behavioural data. Selected properties that are relevant to the course of a social phenomenon, but also too complex for direct modelling, are replaced in simulation by values measured in the real system. In simulation of reputation system presented in chapter 3.1. behavioural data are used to determine when and how often sellers offer new goods. Such approach allows us to greatly increase simulation reality and, at the same time, does not make it too complicated.

[53] Even when we are able to reflect all aspects of the modelled phenomenon we often opt not to do so because of the complicated nature of the phenomenon and the need to build a model which is easier to understand.

2. Sensing social phenomena

2.1. Spiral of hatred: social effects in Internet auctions. Between Informativity and Emotion[54]

2.1.1. Introduction

Transaction volumes and numbers of users in e-commerce systems have been booming over the past few years and there is no sign of a slowdown in the foreseeable future. Every new account in an auction house or within web 2.0 services creates new challenges for privacy and security. Auction houses seem to be the most demanding environment for trust management systems due to direct relationship between reputation and users' income (Lucking-Reiley et al. 2007; Bajari and Hortacsu 1999). Every unpunished and undetected fraud undermines the honest agents' motivation to play fair. Thus, many researchers are working to create new reputation algorithms. Nevertheless, reputation management systems embedded in the most popular websites remain practically unchanged over years and are based on very simple quantitative evaluations and qualitative comments.

Thus far, most researches have been focused on improving algorithms using qualitative feedback and therefore there is a relatively narrow selection of papers devoted to mining comments (security perspective [Gregg, Scott 2008], trustworthiness of reviews [Talwar 2007]) and developing algorithms for trust management systems which explicitly consider descriptive opinions (Ganesan 2008). This is so partly because natural language processing module, which is the cornerstone of such an algorithm, requires building it almost from scratch for every single language (the reusable part is insignificant). It means that the results obtained for languages other than English are hardly comparable and difficult to validate for a larger scientific community.

Nevertheless, it makes good sense to devote resources to the discovery of patterns in descriptive opinions expressed in languages other than English since most Internet transactions are done in language environment native to participants and as local web auction markets grow very fast, this situation will probably continue into the future. Many observations reported in this chapter are likely to apply to other cultures too, irrespective of the language in which the comments are written. Primarily, users' behavioural patterns refer to more general psychological (e.g. spiral of hatred – response is stronger than impulse [Scheff 2004]) and sociological effects which can be even stronger than the cultural fingerprint. A comparative study on Taobao (Chinese version of an online auction marketplace) and eBay has partly confirmed this assumption (Lin 2005).

[54] This research was published as: Nielek, R., A. Wawer, et al. (2010). „Spiral of hatred: social effects in Internet auctions. Between informativity and emotion." Electronic Commerce Research 10 (3): 313–330.

In reference to the above, this chapter is devoted to an analysis of users' behaviour during the after-transaction evaluation process, in particular taking into account pairs of comments on the same transactions delivered by both sellers and buyers (cross-comments) in the auction house. Two approaches have been used to identify and validate different hypotheses. In section 2.1.2., feedback mechanisms existing in e-commerce systems are described. Section 2.1.3. is devoted to quantitative and statistical examination of the collected data and focuses on the effects related to comment type and order in which they arrive. The results obtained by natural language processing algorithms in the context of the hypothesis for validating the spiral of hatred effect are presented in the section 2.1.4. The last section features discussion of the results and a new heuristic model to solve some of the identified problems. The last section presents the conclusions and possible trends in the future research.

2.1.2. Quantitative and qualitative comments

The most commonly used reputation systems embedded in online auction website allow us to evaluate transaction results not only by selecting a predefined category from a list but also by leaving shorter or longer comments. The quantitative measurement in use by eBay and Allegro (the biggest Polish auction house) is based on a very simple structure. When a given transaction is completed, every eBay/Allegro user can evaluate his or her partner by choosing either a positive or neutral, or negative mark. The evaluation mark is visible after being submitted. On eBay it is also possible to evaluate separately the quality of a delivered product, communication, shopping time as well as shipping and handling charges. All those additional evaluation are anonymous. The sums of positive, negative and neutral marks are presented separately. Because feedback is not obligatory, not every transaction is followed by its evaluation. As shown in (Morzy, Wierzbicki 2006) no information is usually indicative of bad experience during the transaction.

Predominantly, only positive comments appear. For more than 1.7 million comments in the collected database there were only ca. 9000 negative and ca. 5000 neutral comments which means that either the fraud level is very low or (it seems more likely) there is a mechanism, which discourages people from making negative comments. Certainly, the threat of legal action (Nam sued for libel over comments on eBay n.d.) constitutes one source of fear, another one is probably related to the possibility of being punished with negative reciprocal evaluation. Yet, another effect identified by researchers (Botsch, Luckner 2008) is that users award a positive quantitative evaluation mark but describe all negative aspects of a transaction in words.

The relative stable framework in the auction houses provides a good opportunity to detect even quite complicated users' behavioural patterns. Abilities of users to learn from previous experiences and to modify their strategies appear to be nontrivial attractors within the space of possible behavioural patterns. A good example of self-adaptation in the complex system which has emerged in online auction websites is that users pay much more attention to negative comments when they calculate transaction risk (Standifird 2001).

Typically, users can intentionally express their opinions only by making comments which are composed of a selected label (quantitative) and a description (qualitative). Nevertheless, a lot of additional information can be found in the data collected in the online auction website, for example response times on positive and negative evaluations, order of buyer-seller evaluation, length of comments or context and reference points (average rating for specific subsets). Identifying measurable effects in buyer-seller interaction can help improve the existing trust management algorithms and create a foundation for designing new ones.

2.1.3. Experiments

2.1.3.1. Dataset

The database analysed in this chapter was provided by the biggest Polish auction house (over 70% of market share). At the beginning of the fourth quarter in 2006, 10,000 sellers and 10,000 buyers have been randomly selected; their profiles and received comments have been stored (description and evaluation). During the next 6 months all transactions conducted by the selected users were monitored and recorded. For every partner who appeared in transaction and was not in the primary database, all historical information about the received feedback has been collected, but with respect to new auctions only the originally selected users have been monitored. In the first quarter in 2007 the database contained more than 200,000 transactions and over 1.7 million comments.

2.1.3.2. Formal definition

Symbols used in the following sections are defined below:

U — set of all users,
T — set of all transaction,
t_m — m-th transaction,
u_i — i-th user,
$c_{t_m}^{u_i}$ — comment left by the i-th user after m-th transaction,
$\tau(c_{t_m}^{u_i})$ — sentiment measured by sentipejd for the comment c,
$\rho(c_{t_m}^{u_i})$ — label for the comment c given by the i-th user,
$r(t_m, u_i)$ — the role of the i-th user in the m-th transaction (either buyer or seller),
$\varphi(c_{t_m}^{u_i})$ — timestamp for the comment c,
$\omega(c_{t_m}^{u_i}) = (c_{t_m}^{u_i}, c_{t_m}^{u_j})$ — an ordered pair of comments for the m-th transaction,
$\delta(\omega t_m)$ — time between two comments.

2.1.3.3. Amount, type, time and order in cross-comments

The objective of this chapter is to identify the effects which appear during bidirectional evaluation, therefore the main focus was an analysis of the ordered pairs of comments, defined in the previous section as ωt_m. For over 1.7 million comments slightly more than 800 thousand pairs were found (in ca. 9% of cases only one party of a given transaction left a comment – either buyer or seller) Only 5056 of pairs contain at least one non-positive evaluation.

Over 90% of answers for comments are made within 14 days after the first evaluation. Shape of curves on the figure 2.1 is similar for all considered cases but there is a notable bias in the starting point. In general, sellers are more responsive – for negative and neutral comments over 20% of sellers and only 7% of buyers feedbacks were written in less than one hour after receiving an evaluation from the partner (for positive comments the numbers are 7% and 3% respectively). On average, buyers seem to visit the auction website less often, so their reaction is slower. Comments, regardless of their contents, are emailed to the evaluated user, thus there is no other variable, except for the type of comment, that may explain the variation in reaction times. Very short response times for negative and neutral comments (when compared with positive feedbacks) can be explained by the will to punish, as fast as possible, the author of the negative[55].

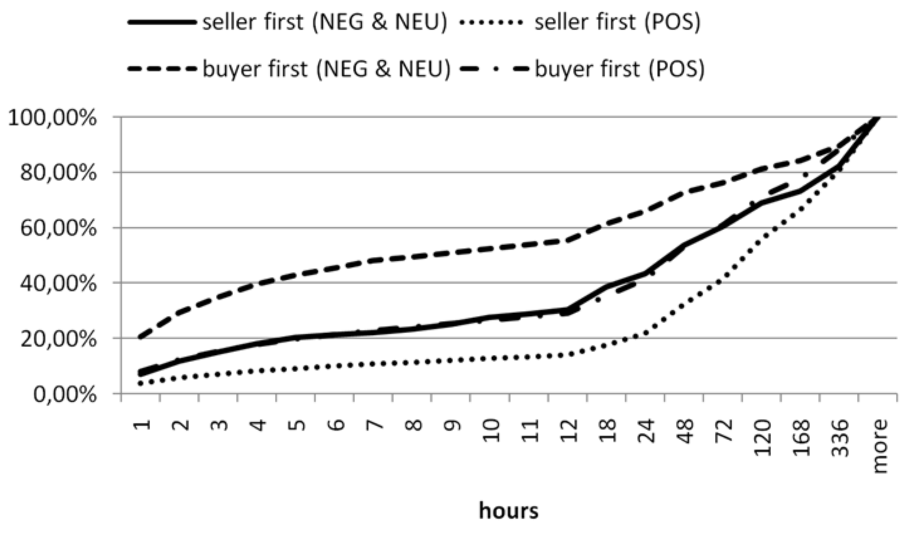


Fig. 2.1. Time-span between comment and answer for different category of evaluation (cumulative histogram)

Average length of comments presented in table 2.1 varies between both transaction roles and different feedback types. As a rule, longer text creates an opportunity to enumerate more facts and express a broader variety of emotions, but also emphasizes the importance of the particular comment for a given user – she or he has

[55] As it is shown in the next sections, neutral comments are very similar to negative comments.

been ready to devote more time to leave feedback. For a positive experience, which is a typically expected result of the transaction, the comments are relatively short – 100 characters in the case of sellers and 73 of buyers (the difference is statistically significant). Manual inspection of a both comment types indicates that the cause of this difference results from the habit of adding an advertisement at the end of comments made by sellers (e.g. "Hope to see you again. ALFRA_PL").

Dissatisfying transaction outcome is positively correlated with length of evaluations. More characters are needed to describe and probably justify dissatisfaction and the negative feedback. The difference between buyer and seller observed for positive comments disappears for both negative and neutral evaluations (small differences observed in table 2.1 in the second and third row are statistically insignificant).

Table 2.1. Average length of comments (in characters)

	Seller	**Buyer**
POS	102.49	73.18
NEU	149.02	154.27
NEG	183.77	178.78

The unwritten rule in online auction websites is that buyers make comments first. For the pairs of comments containing only positive evaluation in 8.2% cases this rule was broken. If one of the comments is negative or neutral, the number of cases contradicting the unwritten law rises dramatically to over 18%. There are many reasons why sellers decide to make a comment first. Some of the sellers probably participate in too many transactions to follow which one is already finished and commented and which not. Strong evidence of such behaviour can be seen in table 2.2 in the very last column – more than 8% of sellers answered to negative evaluation with a positive one. More detailed manual analysis of these pairs indicates that some of the answers contain explanations of the reason for unsatisfactory quality of service (e.g. limited access to the Internet or problems with logistic) but most is given disregarding the previous negative comment. Yet another hypothesis is that the seller is forced by an external event to send the feedback – he or she needs to pay commission for the auction website within a limited period of time after finishing the transaction regardless of its outcome.

Table 2.2. Combination of comment-answer pairs (buyer first)

	POS/x	**NEU/x**	**NEG/x**
x/POS	—	937 (22.60%)	339 (8.17%)
x/NEU	0 (0%)	686 (16.55%)	41 (0.99%)
x/NEG	0 (0%)	408 (9.84%)	1734 (41.83%)

If the buyer is satisfied and expresses this satisfaction with a positive comment first, the answer from the seller will always be positive. All the collected pairs confirm this rule without exceptions. It could be only partially explained by the previous observation. First, a vast domination of positive comments makes a pair $\omega(c_{t_m}^{u_i}) = (c_{t_m}^{u_i}, c_{t_m}^{u_j})$ such as $\rho(c_{t_m}^{u_i}) =$ pos $\wedge\, \rho(c_{t_m}^{u_i}) \neq$ pos statistically very improbable.

Yet, the distribution of such comments was asymmetrical between both transaction roles. On one hand, of over such 550 cases exist for $r(t_{m,} u_i) =$ seller on the other, no such comment pair was found $r(t_{m,} u_i) =$ buyer. Scondly, the results (Standifird 2001) show that even substantial amount of negative feedback does not affect the ability of buyers to participate in transactions. Therefore, a positive opinion about the seller is always rewarded with a reciprocal positive feedback. Thirdly, although there is no explicitly defined procedure to change already submitted feedback, it is essentially possible after some reasonable efforts (e.g. sending an email to the webmaster). So, a seller can refrain from making a non-positive evaluation only because of an aversion to initiating a "war", even though not everything went correct during the transaction.

Table 2.3. Combination of comment-answer pairs (seller first)

	POS/x	NEU/x	NEG/x
x/POS	–	0 (0%)	0 (0%)
x/NEU	284 (31.17%)	19 (2.08%)	19 (2.08%)
x/NEG	242 (26.56%)	8 (0.87%)	339 (37.21%)

Ordered pairs of comments

$$\omega(c_{t_m}^{u_i}) = (c_{t_m}^{u_i}, c_{t_m}^{u_j}) \tag{1}$$

such as

$$r(t_{m,} u_i) = \text{neu} \wedge r(t_{m,} u_j) = \text{neg} \tag{2}$$

(the first evaluation is neutral and the second negative) appear eight times more frequently (i.e. 416) than pairs where

$$r(t_{m,} u_i) = \text{neg} \wedge r(t_{m,} u_j) = \text{neu} \tag{3}$$

(the first evaluation is negative and the second neutral) (i.e. 60).

This enormous disproportion cannot be explained by the course of the transaction because there is no evidence to claim that a negatively affected party will comment second. A more credible explanation is that neutrally evaluated agents use negative evaluations as a punishment and try to do it as severely as possible.

2.1.4. Linguistic productivity

2.1.4.1. Copy-and-paste

The collected dataset contains more than 1.7 million comments but only a fraction of them is unique. Figure 2.2 shows the result of a very simple experiment: for every user category (buyers and sellers) and every feedback type (negative, neutral and positive), distinctive text comments were counted and compared with the number of all collected comments. For the sake of this experiment the strongest possible definition of similarity was applied. Every string was compared character-by-character and only texts which have all characters identical were marked as non-distinctive. The habit identified by this experiment is in fact a copy-paste one.

As can be observed in figure 2.2 all users of the auction house are more willing to devote time to write unique and informative comment when they have experienced either fraudulent (negative) or not perfectly honest behaviour (neutral). Due to the asymmetry of the positive and non-positive comments, virtually all users left only positive evaluations. Hence, they don't have any prepared sentences which they can re-use as an out-of-a-box comment. Because of that, the high percentage of distinct texts for negative and neutral evaluations is an effect of a necessity rather than intentional choice.

Most of the professional sellers who generate the vast majority of transactions in the ebay like auction houses have only one or two versions of positive comments and they re-use them for every non-negative auction. The most popular comment appears 33k times in our dataset (it is slightly less than 4% of all positive comments). Despite the unique buyers to unique sellers ratio, which is more than ten to one in the collected dataset, sellers are even less creative than buyers (5k and 1% respectively). Among the top 50 most popular comments left by buyers over 40 are constructed by combination (different order and punctuation) of only three words: *wszystko* (English "everything"), *ok* (English "ok"), *polecać* (English "recommend").

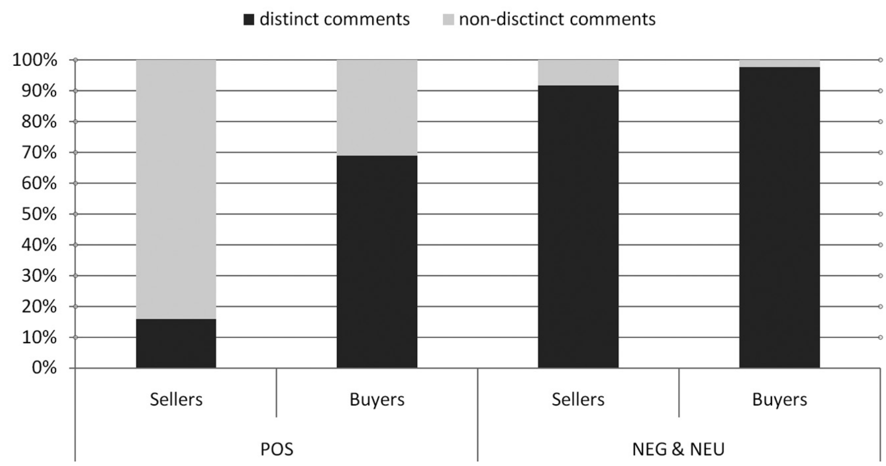


Fig. 2.2. The percentage of the unique comments left by sellers and buyers in the auction house presented separately for the positive and neutral/negative evaluation

2.1.4.2. Linguistic Richness

Research on linguistic richness and language variability involves comparing word frequency distributions following "Large Number of Rare Events" (LNRE) paradigm (Baayen 2001) for computing corpus level type-token statistics. In case of Polish, one can consider token types as either surface word forms or lemmas, each choice having certain advantages over the other. In this text I have followed the second option and assumed token types as baseforms, computed by a morphological analyzer. On one hand, this decision reduced the total number of possible types, but on the other introduced certain overhead due to ambiguous output: one surface form is often resolved to more than one base form.

Insights into linguistic productivity and variability, as used in auction house comments, can be conducted using more than one technique and measure. From the standpoint of information theory, valuable social information is conveyed by productive lingustic processes characterized by relatively high proportion of *hapax legomena*[56] and high probability that adding more texts into the corpus or sample introduces many new, unseen words. The measurement of vocabulary size growth pace is reflected in vocabulary growth curves: relationships between vocabulary size (number of unique words V) and corpus size (number of tokens N). Figure 2.3 presents vocabulary growth curves for following types of comments:

NegNeuBySellers: Negative and neutral comments left by sellers.
NegNeuByBuyers: Negative and neutral comments left by buyers.
PosBySellers: Positive comments left by sellers.
PosByBuyers: Positive comments left by buyers.

The chart reveals two important findings. Firstly, the corpus of positive comments is characterized by much lower vocabulary growth pace than the corpus of negative comments. Secondly, all types of comments written by buyers are more linguistically creative (thus informative) than comments written by sellers. This particular finding can be explained by the fact that buyer-typed comments were left by way more people than sellers-typed comments. Another possible approach is to measure the proportion of token types (word lemmas) that occur only once in the corpus, similar to Harald Baayen's degree of productivity P. High proportion of such words indicates linguistically productive and informative messages. Let P denote linguistic productivity measured as:

$$P = \frac{V}{N} \tag{4}$$

where V denotes the number of word base forms that occur exactly once in a given corpus (subset of comments) and N denotes the total number of all base forms occurrences encountered in comments. Values of P for the four types of comments are presented in table 2.4.

[56] Words which occur only once in the corpus.

Not surprisingly, the highest *P* values have been observed for negative and neutral comments. Relatively high *P* for PosByBuyers indicates that positive comments left by buyers are much more variable than it might seem from figure 2.3.

Table 2.4. P measure for the four types of comments

Comment Type	P
NegNeuBySellers	0.034
NegNeuByBuyers	0.027
PosBySellers	0.007
PosByBuyers	0.021

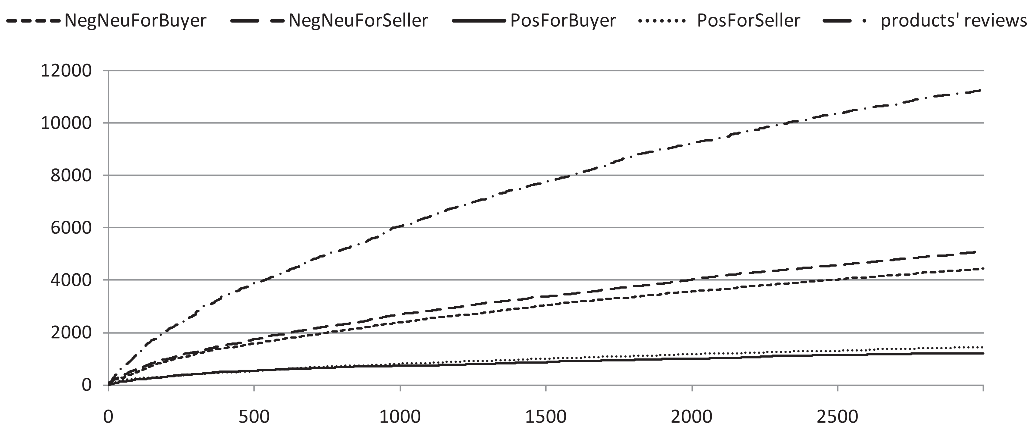


Fig. 2.3. Vocabulary growth curve for four comment types

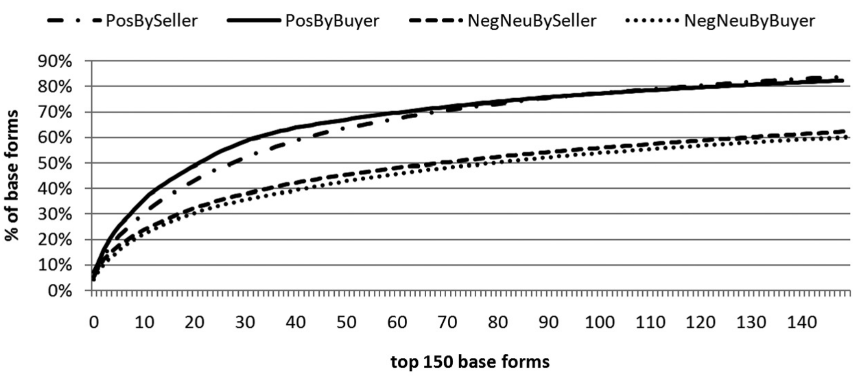


Fig. 2.4. Cumulative distribution of top frequent 150 lexemes for each comment type

As it was already shown in the previous section on the figure 2.3 positive comments are less informative than negative. Cumulative histogram of top frequent 150 lexemes presented on figure 2.4 provides an additional support for that statement.

Over 60% of all words used by buyers in positive transactional experience descriptions adhere to the set of the top 30 base forms. Sellers are only slightly more creative reaching 55% respectively. The same level of coverage of used words can be reached with about 130 top base forms for negative comments. The difference in linguistic creativity between buyers and sellers can be still observed, but is much smaller (ca. 1% for the same amount of base forms).

2.1.5. Mining the meaning of comments

2.1.5.1. Automatic sentiment extraction

For the sentiment analysis task we used a modified version of Sentipejd (Buczynski, Wawer 2008) – a hybrid of lexeme category analysis with a shallow parsing engine. At the basic level, Sentipejd checks for presence of a specific category of lexemes. Such an abstraction originates in content analysis systems, most notably the classic General Inquirer (Philip 1966). Lexical categories used in this work include two sets of words (dictionaries): 1580 positive and 1870 negative ones, created by Zetema[57]. Because comment texts are typed in a careless manner, very often completely without diactrits, lexeme recognition was extended with a diacrit guesser. Recognized sentiment lexemes, along with morphosyntactic tags, are analyzed with Spejd – a tool for simultaneous morphosyntactic disambiguation and shallow parsing (Buczynski 2008), with a number of rules crafted to recognize multiword opinion patterns and apply sentiment modifying operations.

The Spejd formalism is a cascade of regular grammars. Unlike in the case of other shallow parsing formalisms, the rules of the grammar allow for explicit morphosyntactic disambiguation, independently or in connection with structure building statements, which facilitates the task of the shallow parsing of ambiguous and/or erroneous input.

For the purpose of sentiment analysis we extended the default Spejd's morphosyntactic tagset with a sentiment category expressing properties of positive or negative sentiment. This hybrid approach has been called Sentipejd (Buczynski, Wawer 2008).

Sentiment rules, discussed more extensively in (Buczynski, Wawer 2008), included (but were not limited to) the following:

- *affirmation* – an expression of positive sentiment, usually an adverb confirming the sentiment of a positive word and should be treated as strong indications of sentiment (eg. 'I strongly recommend'),
- *negation* – as simple as the difference between "polecam" ('I recommend') and"nie polecam" ('I do not recommend'); the example generic rule captures also statements including the optional verb 'to be' ([base by]), like "nie jest dobry" ('isn't good'),

[57] www.zetema.pl.

– *nullification* – expressing lack of a certain quality or property (usually of negative sentiment), for example "nie mam zastrzeżeń" ('I have no objections'),
– *limitation* – a limiting expression tells us that an expression of positive and neg- ative sentiment has only a very limited extend, therefore hinting that the general sentiment of the review is the opposite of the expression; example: "jedyny problem" ('the only problem'),
– *negative modification* – an adjective of negative sentiment preceeding a positive noun, for example "koszmarna jakość" ('nightmarish quality').

Sentipejd returns either vectors of two integers ($emo_i = [pos, neg]$) which express separately strengths of positive and negative emotions (it's not a simple sum of all emotional phrases) or the single, composite value – $\tau(c_{t_m}^{u_i})$. Every comment present in the collected dataset has been analyzed separately and the result has been stored as a vector in a database together with a category of the comment and the comment itself.

2.1.5.2. Informativeness of Sentiment

Using vectors of two integers [pos, neg] which are returned by Sentipejd, we have introduced a new measure, sentiment informativity, or *SI*, which combines both informativity and sentiment. For this purpose, we have adapted and modified a well-known information retrieval weighting technique, *IDF* (inverse document frequency).

While the classic *IDF*, or inverse document frequency, is typically used with *T F* (term frequency) component in *T F IDF*. Because comment length is limited to 256 characters, it is very unlikely that informative words are repeated in the same comment. Thus, we have omitted the *T F* part. For each token *i*, we are computing its *IDF* score as:

$$IDF_i = log\left(\frac{|D|}{|\{d_i : t \in d\}|}\right) \tag{5}$$

where $|D|$ is the total number of comments, $|\{d_i{:}t{\in}d\}|$ is the number of documents where the term *i* appears. Next, we aggregate comment-level *IDF* scores for "sentimental" tokens and expressions, obtaining two non-negative numbers of positive and negative SI (sentimental informativity), SI_{POS} and SI_{NEG}.

$$SI_{[POS,NEG]} = \sum_{i=1}^{N} IDF_i : i \in [POS, NEG] \tag{6}$$

2.1.5.3. Reclassification precision and the emotional distance

Although a similar natural language processing module has been already applied by authors to a broad variety of subjects (e.g. dynamic of public opinion [Wawer, Nielek 2008]) the very first question which arises is: can a NLP system extract and evaluate emotions from usually very short and not always correctly (grammatical mistakes and typos) written comments? To answer this question, which is crucial for further deliberations, a standard data mining approach was used.

Four separated, balanced subsets of comments were created:

- Set I (POS; NEG) – contains 2590 comments whereof 1295 are negative and 1295 positive,
- Set II (NEG; NEU) – contains 1454 comments whereof 727 are negative and 727 neutral,
- Set III (NEU; POS) – contains 1454 comments whereof 727 are neutral and 727 positive,
- Set IV (POS; NEU; NEG) – contains 2181 comments whereof 727 are positive,
- 727 negative and 727 neutral.

Every set of comments has been partitioned on testing and training set (30% and 70% cases respectively). For every set of comments three different classification approaches were used: neural network, support vector machine and decision trees (CHAID algorithm). As a target variable the label given by comment's author (negative, neutral or positive) was selected and the emo vector as input variables.

Table 2.5. Classification accuracy for different algorithms and testing subsets (average of four runs)

	Neural Network	**Support Vector Machine**	**Decision Trees (CHAID)**
POS and NEG (two classes; set I)	90.86%	90.36%	89.37%
POS, NEU and NEG (three classes; set IV)	61.58%	61.66%	61.79%
NEG and NEU (two classes; set II)	65.16%	65.80%	64.11%
NEU and POS (two classes; set III)	71.15%	69.15%	70.24%

The obtained results are presented in table 2.5. The first experiment was conducted to check if an evaluation based on the emotions expressed in comments and measured by the Sentipejd allows to predict the polarity of an label given by a human. At the beginning, the simplest subset was tested (only two classes – positive and negative – which should be relatively easier to separate). For the first set neural network approach was the most efficient. Over 90% classification accuracy indicates that the Sentipejd deals quite well with extracting emotions from texts (even not 100% correctly written) and that the significant difference in emotional content between positive and the negative labelled comments can be confirmed and measured.

Similar results for the neural network, support vector machines and decision trees (90.86%, 90.36% and 89.37% respectively) suggest that the reason for wrong classification goes beyond the classification algorithms. Only slightly better results for the same algorithms but validated on training sets instead of test sets seem to confirm that as well. A closer look at the misclassified cases shows that they belong into three (not always distinct) groups:

- written in a very specific slang, many misspellings, grammatical and orthographical errors, a lot of emoticons,
- well written but based on ironic, quizzical description of the past transaction,
- marked by user as positive but containing a negative evaluation.

The existence of the third group seems to confirm the results presented by Botsch and Luckner (Botsch, Luckner 2008). Some users, instead of leaving a negative mark, prefer to describe all the experienced problems in words. Because of their incoherency, those cases cannot be correctly classified using the adopted approach and they should be removed from the database. A detailed estimation of the scale of this effect requires manual processing of every comment which is not feasible because of the database size (1.7 million comments) and extends beyond the scope of this chapter, although a rough estimation indicates that the effect of incoherent feedbacks is lower than 0.1% of all positive comments.

The biggest fraction of wrongly classified comments belongs to the first group. Many users, not only in e-Commerce systems but also on online forums, use a lot of abbreviations, emoticons, colloquial words and even intentionally misspelled words. Frequently, using intentionally transformed words is a sign of being a member of a specific social group. It helps users to identify the newcomers in an environment where cheap pseudonyms are present (a detailed study of the effects introduced by using cheap pseudonyms can be found here [Friedman, Resnick 2001]). Some problems can be resolved (e.g. using a spell-checker to correct orthographical mistakes or creating a dedicated dictionary containing slang and colloquial words) but in principle intentional modifications of meaning or detecting irony will always be a challenge for computational linguistics.

The results for set IV are presented in the second row in table 2.5. Introduction of the third class made the task much more difficult. The results over 60% are still almost 30% better than in the baseline of random choice but significantly lower than for two classes. Thus, to check which comments cause problems for the classification algorithms, two more experiments have been conducted. Firstly, the separability for neutral and negative comments has been tested. The third row in table 2.5 contains the results for the set II which includes only negative and neutral comments. The classification precision slightly over 65% indicates that the emotional distance between neutrally and negatively tagged feedback is relatively small. Secondly, the same approach has been used to measure the emotional distance between neutral and positive comments. The results for all classification methods except support vector machines are at least 6% better and indicate that neutral comments are emotionally closer to negative.

To confirm the hypothesis stated in the previous paragraph a new testing set has been created. All the collected comments were split into two classes: one containing only positive labelled comments and one with negative and neutral feedback. Based on the emo vector (defined at the beginning of this section) and using the classification algorithms (support vector machines, neural network and decision trees) an attempt to rediscover the new classification has been done. The obtained results are slightly less precise than for the set I (positive and negative comments only; without neutrals) but the difference is about 3%. Thus, in most applications negative and neutral comments can be interpreted in the same way – as an expression of dissatisfaction. The label should not be treated as a scale of the experiences) because there is very little data to confirm the hypothesis that neutral feedback is less effective than negative.

2.1.5.4. Spiral of hatred

The spiral of hatred is a well-known phenomenon present in a wide area of scientific fields (eg. Wydra identifies it as an core component of the war conflicts [Wydra 2008]) manifested as an endless action-reaction response, where successive iterations are subject to more negative emotion. Typically, in practical terms this effect can be observed on online forums where an initial misunderstanding causes a lasting exchange of messages containing many abusive words. Because reputation influences profitability of the seller (Standifird 2001) and every negative comment undermines this reputation, thus the reaction of a seller after receiving negative feedback can be more emotional. In fact, as a consequence of unbalanced levels of positive and negative comments, a very interesting heuristic has emerged. For experienced users, a single negative comment plays much more important role in the estimation of transaction risk that even many positive comments.

Because comments are visible after they are left, a natural place to express (and observe), the spiral-of-hatred effect is the reciprocal feedback given by the second party after transaction is completed. Thus, the database described in section 3.1. has been used to verify the spiral-of-hatred effect, which – referring to the formalism defined in section 2.1.3.2. – can be expressed as:

$$\forall \rho\left(c_{t_m}^{u_i}\right), \rho\left(c_{t_m}^{u_j}\right) \in \{NEG, NEU\} : \varphi\left(c_{t_m}^{u_i}\right) < \varphi\left(c_{t_m}^{u_j}\right) \rightarrow \tau\left(c_{t_m}^{u_i}\right) > \tau\left(c_{t_m}^{u_j}\right) \tag{7}$$

It is reasonable to assume (because of the sociological nature of the analyzed effect) that the above definition will not apply universally and to every single case. Therefore, in the first stage a weaker assumption was tested – the average of negative emotion for the second comment is higher than for the first:

$$\forall w, u \in U; s, t \in T : \left(p\left(c_s^u\right), p\left(c_t^w\right) \neq POS : \varphi\left(c_t^w\right) < \varphi\left(c_s^u\right)\right) \rightarrow \sum_{w,t} \tau\left(c_t^w\right) < \sum_{u,s} \tau\left(c_s^u\right) \tag{8}$$

The results are equivocal. First, the average value of τ for the comments given first is –0.63. The same value for the answers is higher and amounts –0.72. The difference is statistically significant at the level 0.07 which is a little bit above a typical 0.05 but it seems to make a spiral of hatred hypotheses at least very probable. Second, the standard deviations for both sets are almost equal –2.01 – and it indicates that the distribution of emotion intensity between both the earlier and latter comment groups is similar but shifted. On the other hand, dividing the set analyzed in the previous paragraph into buyer and seller roles of the agent, makes results more complicated. More detailed results are presented in table 2.6.

In general, sellers are more emotional and more expressive than buyers (τ = –0.76 as compared to τ = –0.60 for buyers) and this pattern concerns both specific cases analyzed in table 2.2. There are at least three hypotheses which can explain this difference. Firstly, sellers write more correctly so the Sentipejd has an easier job extracting emotions from comments. Secondly, the cost of receiving negative evaluations for sellers is much higher (pseudonyms are more expensive and lower reputation affects profitability) and therefore boosts their reaction. Thirdly, sellers are simply more experienced and know how to make comments in a more negative way. As the standard deviation for sellers is only slightly higher than for buyers (should be significantly higher, if the source of difference in the emotional strength is related to misspellings and errors in comments), the second and the third hypotheses are the more probable ones.

Table 2.6. Average sentiment for buyers and sellers

	First	Second
Buyer	0.55	–0.69
Seller	1.28	–0.96

The classical definition of the spiral-of-hatred effect formalized in eq. 7 and 8 is satisfied (and statistically significant) only for typical cases where buyers leave comments first. As expected, the average answer given by a seller is more negative. However, the same assumption is not true for the uncommon situation where sellers comment first. In that case the average negative sentiment in ordered pair $\omega t_m = (c_{t_m}^{u_i}, c_{t_m}^{u_j})$ such as $r(c_{t_m}^{u_i}) = seller$ is –1.12 and is much higher (–0.96) than for $r(c_{t_m}^{u_i}) = buyer$. The increase in negative emotions, compare to situation where buyers comment first, is observed symmetrically for both participants (buyers and sellers). Even though buyers answer very aggressively, at the end the emotional war is always won by sellers. They have stronger motivation because the reputation affects their profitability and are more experienced due to the extensive usage of the auction website.

More studies are needed to determine how the communication beyond auction platforms' cross-comments mechanism (e.g. via e-mail) influences emotional attitudes. However, on the very basic level the spiral-of-hatred effect can be identified in the collected data despite complex interactions of many social processes.

2.1.6. Discussion

Originally, the auction houses have been developed as goods exchange platforms where everyone could be either a seller or a buyer and where such roles are volatile and adopted only for one transaction. Nowadays, the auction platforms remind more of a shopping mall rather than a medieval bazaar and almost all members have clearly defined typical roles of either sellers or buyers. Therefore, it is necessary to revise the previous paradigm which used to determine the development of the reputation management systems. Instead of two more or less equal transaction parties, there is an explicit distinction: on one hand, sellers become more experienced due to the extensive usage of the auction system, on the other hand buyers' profitability is less sensitive to negative feedback.

The modification of the reputation system should take into consideration these facts. One of the possible ways to take them into account is for example to limit the possibility of leaving an evaluation by making it available only for buyers. Onesided comments make sellers defenceless, but elimination of negative reciprocal feedback will increase the likelihood that buyers comment more honestly. As an undesirable side effect of such a situation, blacklists of dishonest buyers can be created and maintained outside auction platforms, which can in turn be used as a tool for sometimes unjustified discrimination. More side effects should also be expected.

Another way to eradicate the spiral-of-hatred effect, which requires merely a minor modification in the existing reputation management systems, is to hold back the publication of an evaluation until an answer is sent. It should permit the elimination of the threat of revenge and thus make all comments more honest and less biased by the previous evaluation (more an answer based only on transaction experiences than an evaluation of the other participant performance). The problem that users will intentionally block publication of the negative comments can be solved by introducing a moderator who will be responsible for making an opinion visible (upon a request of one of the transaction parties) even if the answer does not appear. Even fewer changes are required to reduce the identified effect through establishing the minimum time-span that has to pass between comment and answer. Answers given right after a negative comment is received are more emotional and usually less informative.

Natural language processing tools are the best solution to investigate problems referred to in the descriptive part of submitted comments. Automatic sentiment extraction helps identify emotional wars immediately after they appear and either inform the administrators or even take appropriate steps automatically. Analysis of every pair of comments can be complemented by the knowledge about typical behaviour of users taking part in transaction on the basis of their previous evaluations. Moreover, an efficient NLP algorithm can detect many discrimination strategies such as using a multitude of fake pseudonyms or atypical positive evaluations.

2.1.7. Conclusion

The broad variety of effects identified and described in this chapter is only a fraction of all effects in auction websites. Jointly, with the stoning, slipping, self-selection (Khopkar, Li and Resnick 2005), cheap pseudonyms (Friedman, Resnick 2001), asymmetrical impact of positive and negative comments (Standifird 2001), price-reputation correlation (Lee, Im and Lee 2000), the importance of missing feedback (Morzy, Wierzbicki 2006), the presented results provide environment for invention, development and implementation of new techniques and tools with a goal to further increase satisfaction and usability of an auction website. Proposed changes can impact not only users' satisfaction but also profitability of the auction website.

The complex relationships between different users' behavioural patterns and hardly predictable side-effects discourage the managers responsible for maintaining and developing e-commerce systems from modifying the existing, proved solutions. They tend to use simple financial instruments like insurances or escrows to increase the level of security. Thus, the attempts to popularise the results collected by researchers over the last few years should be focused on the development of dedicated external tools to support users using those systems rather than on the modification of existing e-commerce systems.

Future research should be oriented toward sensitivity analysis of identified effects and influence of cultural circles and individual characteristics on the dynamics and existence of particular effects. Also, forecasting of social acceptance and social effects of the planned changes in an auction house is a challenging task (Kleinberg 2008). Successful modeling and forecasting social responses (i.e. emergent attractors, stability points, non-linear dynamics) will be crucial to implement changes in Web 2.0 services.

2.2. Sentiment and the Polish stock market. Towards automated financial web mining[58]

2.2.1. Introduction

For every trader the most valuable good is information. Privileged access to information about company or macroeconomic environment creates advantage which can be measured in terms of money. The winner is always someone who first gains information. The importance of information is stressed on almost every stock market by special law which prohibits inside trading (using confidential or inside information when making investment decisions). If the statement written down at the

[58] This research was published as: Wawer A., Nielek R., *Sentiment and the Polish stock market. Towards automated financial web mining*. In Polish Journal of Environmental Studies, vol. 18, No. 3B, 2009, pp. 282–285.

beginning of this paragraph is really true then all the emotions embedded in all the publicly accessible information sources should be able to explain the stock market behaviors. Thus, the aim of that research was an attempt to verify this hypothesis for the Polish stock market based on the Internet as a source of information and NLP tool for automatic sentiment extraction.

In the following section a review of the existing approaches and working systems are presented. Data collection procedure and created ontology are sketched in section 2.2.3. The natural language processing tool used for sentiment extraction is described in the next section. In the last section the results are presented and discussed as well as the key factors which have influenced the Warsaw stock market are named.,

2.2.2. Review of existing approaches

Combining stock exchange data with processing opinion-forming texts, published in media, is a well-known research problem.

Most works and systems constructed them are geared towards optimizing investor strategy and purchase decisions, either on stock or index level. An interesting overview of various prototypes created during recent years can be found in (Mittermayer, Knolmayer 2006). All described systems have a hybrid character, in that they combine a web mining component, natural language processing (using text mining, information retrieval, text classification or any other approach), data processing (machine learning and classification) and decision-supporting (expert and rule-based systems). Each of the systems is built around different principles and different designs. For example, text categorization modules have been created using:

– a dictionary of 423 hand selected features, put into pairs with an AND operator (Cho 1999),
– automated feature selection with TF IDF method (Lavrenko et al. 2000),
– lists of expressions, involving logical operators and string matching, where occurrence of at least one expression in text resulted in assigning it into one of 39 categories (Seo, Giampapa and Sycara 2002).

The list of possible designs is far from complete and reveals important methodological and design differences between described components. Each of the mentioned approaches is essentially a variant of text categorization, which can be replaced by information retrieval methods and keyword-level (typically company name) analysis.

Ambiguous benchmarking criteria and unclear methods of evaluation are a serious problem in case of discussed systems. Each of the applications mentioned in (Mittermayer, Knolmayer 2006) has been created using different sets of texts, data from different stock exchange markets and different periods. The most common evaluation criteria applied by authors of the systems to evaluate quality of own work are profits from transactions suggested by their systems (over certain period) and effectiveness in predicting changing trends.

2.2.3. Automated extraction from text

The approach to automated sentiment analysis from text is similar to (Buczynski, Wawer 2008), with one main difference: namely, the analysis presented in this chapter was done not on text (message) level, but instead was measured towards keywords (names of companies present in WIG 20) in selected chunks of text. Chunks were computed by a purposely crafted algorithm, which took into account syntactic information and selected only those phrases, that were the most likely to be affecting given keywords.

Examples of fragments evaluated by the system as negative(–), positive(+) or netural(=):

„Stalexport wygrał przetarg" / „Stalexport won the tender" (+),

„Największy skok zysku spodziewany jest w przypadku Budimeksu" /

„The biggest jump of revenue is expected in the case of Budimeks" (+),

„kurs Sygnity jest najniższy od 12 lat" / „Sygnity's stock rate is at a 12 years low" (–),

„Strata Sygnity coraz większa" / „Sygnity's loss gets bigger" (–),

„W Kredyt Banku mają teraz ofertę Ekstrakonto" /

„In Kredyt Bank, they have now offer for Ekstrakonto" (=).

This approach targets increased precision, but manual analysis of one data fragment indicated that in about 1/3 of cases it was not sufficient to ascertain, what the attitude towards a given keyword is. In such cases one had to read the whole message, going beyond the selected chunk (text fragment). The problem can be addressed only after introducing initial text classification as to their structure, both linguistic and semantic, to distinguish texts where chunk-level is sufficient from those, where message-level analysis is necessary. This solution involves two parallel streams of processing, according to recognized content type.

2.2.4. Data collection

The texts analyzed in this chapter were downloaded daily between 17th of September and 7th of November 2009 from the 20 main Polish news portals (section economy and finance) and forums where the hottest discussion are conducted by day traders. Crawler using the focused-crawling technology (Chakrabarti 1999; Lawrence, Giles 1998; Rungsawang, Angkawattanawit 2005), travelled through the web in a recurrent manner with depth set to 2. Every day more than 18.000 pages were visited and downloaded. From downloaded pages only the content of the articles has been extracted and analyzed. Navigation bars, static parts of the pages and ads were filtered out. The tracked set of keywords consisted of twenty distinctive groups of companies which belong to WIG20 index. For each company many different versions of name spelling were taken into account (trivial names, abbreviations etc.).

2.2.5. Macroeconomic environment

The time span which was studied in this chapter belongs to the toughest economic downswing in the 20th century. Paul Krugman, the Nobel prize winner, calls the present situation depression and claims that the well-known solutions, similar to those used during the Great Depression in the 1930s, e.g. cutting the internal rates, will not work[59]. Over the last year the stock markets were (and still are) ruled by bulls. Long and overwhelming market slips were mixed with short and fierce burses of optimism. Volatility Index (VIX) proposed in 1998 by Robert E. Whaley has grown to the level unseen from the beginning of 1970. It shows that even a small piece of information can turn the sentiment on the market upside down. Many well-known formations used for technical analysis like head and shoulders and bearish diamond which usually needed a week or two to develop have been speeded up by unusual volatility and can be observed within one session. Neither Polish nor Western European nor US markets are free of that effect. Not only stock traders but also journalists and analysts swing from bullish to bearish mood.

Far from the stock markets, in the real economy positive information actually does not exist at all. The Euro zone, Japan and the USA went into recession. The unemployment rate is growing rapidly. Over the last 6 months at least four countries went de facto bankrupt – Iceland[60], Ukraine[61], Latvia and Hungary[62]. Abruptly decreasing GDP in many countries indicates that we can expect more states will request assistance of IMF and EU. Concurrently, even the biggest private companies are asking governments around the world for loans and recapitalization. The US government has been forced to spend 170 billion USD to nationalize AIG – one of the most known insurance company. The German government took similar steps in the case of Hypo Real Estate, Belgian concerning Fortis Group and England regarding Lloyds. Such actions didn't calm the stock market, on the contrary it was like adding fuel to the fire. Some of the state authorities had to ban trading on local markets.

[59] http://www.newsweek.com/id/171871?tid=relatedcl

[60] The emergence loan granted by Scandinavian countries and Poland together with the nationalization of the bank sector and prohibition of payouts from savings accounts allowed Iceland to survive the last few months.

[61] International Monetary Fund has granted 16 billion USD loan (because of political instability on Ukraine only the first part has been transferred).

[62] Without the rescue plan conducted jointly by the EU and IMF which allow Hungary and Latvia using line of credit worth 25 billion EURO both countries would have to declare bankruptcy.

Similarly as other stock markets, the Warsaw Stock Exchange (WSE) was governed by bulls crawling sometimes more than 70% below the one year maximum but it avoided suspension. Very aggressive trading of foreign investors on the derivates market combined with a significant decline in the daily trade volume (400 million PLN as compared to over 2 billion PLN a year ago) created an opportunity to manipulate the prices on the stock exchange. Additionally, over the last few months the Polish currency has continued this downward trend and stopped slightly above the historical minimum.

2.2.6. Data analysis

In order to present the relationship between number of positive fragments about a given keyword K (company name or its variant), denoted as pos(K) and a number of negative fragments neg(K), a new index E has been introduced, defined as follows:

```
if (pos(K)>=neg(K)) E = pos(K) / neg(K)
else
if (neg(K)>pos(K)) E = - neg(K) / pos(K)
```

The chart below shows daily index E and WIG 20 values, normalized by dividing previous day's data by current day's data. This normalization has been applied to both data series, index E and WIG 20.

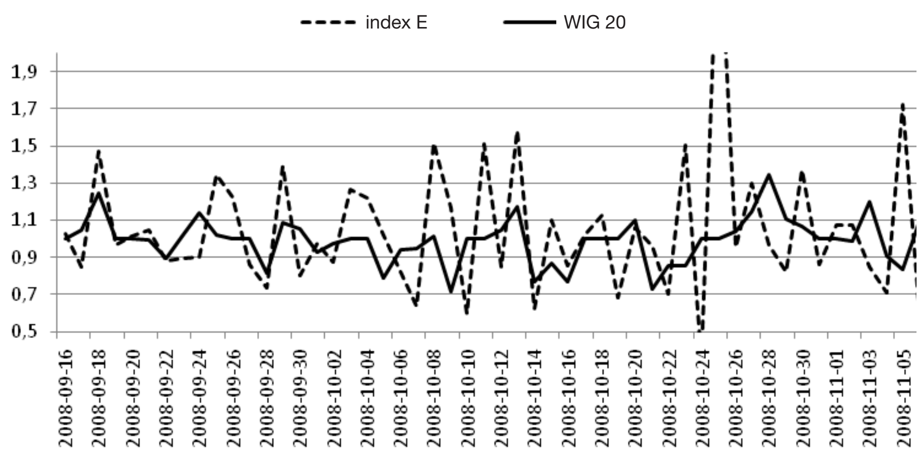


Fig. 2.5. WIG20 and index E over time

Certain similarity of both series can be assessed visually, even without mathematical tools. The most interesting situation for investing strategies occurs when index E changes direction prior to the WIG 20 – this took place on 26th September, 24th and 26th October, 1st of November. In many cases the opposite situation could be observed: changes in the WIG 20 were reflected by the next day's changes of index E. This raises questions about the possible causal relationship between texts, published in the Internet, that contain certain evaluations towards company names, and their influence on WIG 20. Some similarity between the texts can be observed using Pearson correlation, where r value is equal to 0,65[63].

In order to verify a possible causal relationship between both variables, namely index E and WIG 20, we have tested it using Granger's causality, using index E data as a possible Granger-cause of WIG 20. The test revealed delayed index E parameters as statistically insignificant, thus one cannot reject the null hypothesis, which states that index E is not a Granger-cause of WIG 20. Obviously, this conclusion can be supported by visually examining the chart presented above, where index E is only in rare cases ahead of the WIG 20.

2.2.7. Conclusion

This chapter explores relationships between negative/positive sentiment toward names of Polish companies (listed on the Warsaw Stock Exchange) extracted from texts appearing in the Polish Internet and the WIG20 index. Certain level of similarity has been identified but a causal relationship between emotional index and the stock exchange cannot be confirmed. Thus, it is not possible to create a profitable investment strategy based only on collected and processed data. Additional knowledge is required.

Forecasting and prediction of daily changes on financial markets (especial on the stock market) based only on texts collected in Internet is much more difficult than we had initially expected and cannot be compared to prediction of the level of public support for political parties which we has been already done successfully (Wawer and Nielek 2008). The changes of public opinion are driven almost exclusively by media coverage and thus the predictive models for politics can be quite simple. The dynamic of financial markets is influenced by many internal and external factors which seem to be irreducible to either positive or negative sentiment expressed in news. However, the proposed method can be applied as one of several inputs of a bigger system, supporting investor decisions basing on various data, like foreign stock indexes and information from different types of markets.

[63] Significance = 0,00001.

3. Influencing social phenomena

3.1. Fairness Emergence in Reputation Systems[64]

3.1.1. Introduction

In distributed, open systems (ODS), where the behavior of autonomous agents is uncertain and can affect other agents' welfare, fairness of resource or cost distribution is an important requirement. An example of such situations are Peer-to-Peer systems that rely on the sharing of peer's resources. Unfair distribution of the provided resources can occur if some peers free-ride on others. While some systems (like Bittorrent) combat free-riders, they usually cannot achieve fair distributions of resources. In particular, peers who have provided many resources in the past may not receive a similar amount of resources when they need them. Another example are grid systems, where the scheduling of tasks should take into account the fair distribution of available computational resources.

Assuring fairness of resource distributions in a system without centralized control is difficult. On the other hand, in such systems, trust management (TM) is widely used. Examples of practical use of trust management are (among others) reputation systems in online auctions and Peer-to-Peer file sharing systems. Trust management is aimed to provide procedural fairness: to ensure that peers who violate rules or norms of behaviour are punished.

The question considered in this chapter is whether or not TM systems can also be used to assure or increase fairness of resource distribution. While this is different (and more difficult) from procedural fairness, the two concepts are related. Norms and rules of behavior are often defined with the fairness of resource or cost distribution in mind. As an example, consider the laws that oblige all citizens to pay taxes. Enforcing procedural fairness (abiding by the tax laws) has the goal of enabling efficient resource redistribution by the government (among its other duties). Such resource redistribution should result in increased fairness of income distribution.

The question considered in this text can be formulated as the following hypothesis: in successful reputation (or trust management) systems, fairness should be an emergent property[65]. We shall refer to this hypothesis as the Fairness Emergence (FE) hypothesis. In this dissertation, the FE hypothesis has been verified.

We will use a simulation approach to verify the FE hypothesis. However, our goal is not just to see whether the hypothesis applies in an abstract model, but

[64] This research was published as: Wierzbicki A., R. Nielek, Fairness Emergence In Reputation Systems, "Journal of Artificial Societies and Social Simulation" 2011, No. 14 (1) 3.

[65] By emergence we understand the arising of a complex property (fairness) out of simpler system behavior (the use of a reputation system by agents).

to verify the validity of the FE hypothesis in realistic conditions. In order to realize this goal, we need to study the behavior of a popular, well understood trust management system. The natural candidate for such a system is the reputation system used by Internet auctions. Previous studies have established that the use of reputation systems increases the total utility of agents (Pollock 1992; Resnick, Zeckhauser 2002), and investigated the sensitivity or reputation systems to selfish or malicious user behaviour (Dellarocas 2000). This study investigates how the use of reputation impacts the fairness of the distribution of agents' utilities.

The goal of verifying the FE hypothesis under realistic conditions can be fulfilled by a study of Internet auction systems under non-stationary conditions, and in the presence of selfish and malicious users. Reputation systems used in other applications, such as P2P networks, are vulnerable to the same effects. Therefore, our model of a reputation system is sufficiently general to apply to different applications, while at the same time we are able to draw on the well-known properties of reputation systems used in Internet auctions in order to increase the realism of our model. This is done at the risk of drawing conclusions that will apply mostly to Internet auctions. The realism of our study of reputation systems for Internet auctions is increased further by the use of trace-driven simulation (to our knowledge, this is the first such study described in the literature. We have obtained a large trace from a Polish Internet auction provider that is used in the second group of simulations to realistically model agent presence in the system. However, the results from our first group of simulations are sufficiently general to warrant drawing conclusions about the FE hypothesis in other applications domains.

Also, the fairness of distributions of users' utilities in Internet auctions is an important goal in its own right. Buyers or sellers in Internet auctions expect that if they behave as fairly as their competitors, they should have a similarly high reputation. In other words, the users of a reputation system expect that the reputation system give a fair distribution of reputations. In the absence of other differentiating factors, this should also ensure a fair distribution of utilities. This expectation of users is a consequence of the general social norm: people expect fair treatment from many social and business institutions, like a stock exchange, or an Internet auction site.

The questions considered in this work are therefore the following: is the FE hypothesis universally true? Does the FE hypothesis apply under realistic conditions? How sensitive is fairness emergence to the performance of a TM system? What are the conditions that can lead to a lack of fairness emergence due to the use of a TM system? Does fairness emergence occur if agents are infrequently unfair? Does fairness emergence occur if agents have a low sensitivity of to reputation? Does fairness emergence occur if agents employ discrimination? These and like questions can lead to a better understanding of the ability of trust management systems to increase fairness of distribution of costs or resources in an open, distributed system without central control.

In order to evaluate distributional fairness, it becomes necessary to define it precisely. In this work, fairness is defined based on a strong theoretical foundation: the theory of equity (Kostreva, Ogryczak 1999; Kostreva, Ogryczak and Wierzbicki 2004). The concept and criteria of fairness in trust management systems, based on the theory of equity, are discussed in the next section. Section 3.1.2. also discusses the chosen approach to test the FE hypothesis by laboratory evaluation of reputation systems. Section 3.1.3. describes the simulator used for verifying the FE hypothesis, along with the trace-driven simulation approach which is a major contribution of this work (to our knowledge, it is the first trace-driven simulation of a reputation system of an Internet auction site). Section 3.1.4. describes the results of a simpler experiment with a closed system, and the sensitivity of fairness emergence to various aspects of a reputation system. Section 3.1.5. describes the results of the trace-driven simulation that partially support the FE hypothesis. Section 3.1.6. concludes the chapter.

3.1.2. Related work

Reputation systems have usually been studied and evaluated using the utilitarian paradigm that originates from research on the Prisoner's Dilemma. Following the work of Axelrod (Axelrod 2006), a large body of research has considered the emergence of cooperation. The introduction of reputation has been demonstrated as helpful to the emergence of cooperation[66]. In the Prisoner's Dilemma, the sum of payoffs of two agents is highest when both agents cooperate. This fact makes it possible to use the sum of payoffs as a measure of cooperation in the iterated Prisoner's Dilemma. This method is an utilitarian approach to the evaluation of reputation systems (Mui 2003; Dellarocas 2000; Wierzbicki 2006). In most research, a reputation system is therefore considered successful when the sum of utilities of all agents in the distributed system is highest. Note that the utilitarian paradigm is used even if the simulation uses a more complex model of agent interaction than the Prisoner's Dilemma.

The use of Prisoner's Dilemma allows for an implicit consideration of agent fairness, while the sum of utilities is considered explicitly. Yet, in a more realistic setting, the assumptions of the Prisoner's Dilemma may not be satisfied, and it is possible to point out cases when the utilitarian approach fails to ensure fairness: in an online auction system, a minority of agents can be constantly cheated, while the sum of utilities remains high. A notable example of explicit consideration for fairness of reputation systems is the work of Dellarocas (Dellarocas 2000). An attempt to demonstrate that explicit consideration of fairness leads to different results in the design and evaluation of reputation systems has been made in (Wierzbicki 2007).

[66] However, note that the existence of reputation information is a modification of the original Prisoner's Dilemma. Axelrod has explicitly ruled out the existence of reputation information in his definition of the game.

3.1.2.1. Distributional Fairness and the Theory of Equity

Much of the research on fairness has been done in the area of the social sciences, especially social psychology. The results of this research allow to understand what are the preference of people regarding fairness, and how people understand fair behaviour. Interestingly, much of the research in that area has been influenced by the seminal work of Deutsch, who is also an author of one of the basic psychological theories of trust (Deutsch 1975; Deutsch 1987). To begin our discussion of fairness, let us begin with three general kinds of fairness judgements identified by social psychology (Tyler 1998): *distributive fairness*, *procedural fairness* and *retributive fairness*. Distributive fairness is usually related to the question of distribution of some goods, resources or costs, be it kidneys for transplantation, parliament mandates, or the costs of water and electricity. The goal of distributive fairness is to find a distribution of goods that is perceived as fair by concerned agents. Procedural fairness focuses on the perceived fairness of procedures leading to outcomes, while retributive fairness is concerned with rule violation and the severity of sanctions for norm-breaking behaviour. It is possible to think of distributive fairness as a special kind of procedural fairness. If a distribution problem can be solved fairly, then a fair procedure would require all agents to take a fair share of the distributed good or cost. Procedural fairness, however, is also applied in the case when a fair solution cannot be found beforehand or cannot be agreed upon. Both distributional fairness and procedural fairness aim to find fair solutions of distribution problems.

The most abstract definition of fairness used in this chapter is therefore as follows. *Fairness* means the *satisfaction of justified expectations of agents that participate in the system, according to rules that apply in a specific context based on reason and precedent*[67]. This general definition applies to distributive, procedural or retributive fairness. However, for the purpose of testing the Fairness Emergence hypothesis, we have decided to use the concept of distributive fairness, since people care most about the fairness of outcomes, not procedures (distributive fairness is a concept closely related to social justice (Rawls 1971). Although extensively studied (Young 1994), distributive fairness is a complex concept that depends much on cultural values, precedents, and the *context* of the problem. Therefore, a *precise and computationally tractable definition* is needed to use it in research.

The understanding of the concept of distributional fairness in this chapter is based on the *theory of equity*[68] as presented in (Kostreva, Ogryczak 1999; Kostreva, Ogryczak and Wierzbicki 2004). Before we introduce the theory formally, let us attempt to give a more intuitive understanding.

[67] According to the Oxford English Dictionary, the word „fair" means: equitably, honestly, impartially, justly; according to rule.

[68] The term theory of equity may be applied to various axiomatizations of equity described in the literature (Feurbaey 2008), as well as to work in ethics following the seminal work of Rawls. In this texy, I apply this term to a simple axiomatization that has a direct relation to the fairness criteria used in the Generalized Lorenz curve.

In a fair distribution problem, all agents' outcomes must be taken into consideration. The problem of optimizing the outcomes of all agents can be formulated as a multicriteria problem. We shall refer to this as the *efficient optimization problem*. Efficient optimization of agent's outcomes need not have any concern for fairness. The outcomes can be the shares of goods or costs received by agents in an ODS. Let y = [y1,…,yn] be an outcome vector of the efficient optimization problem (assuming there are *n* agents that maximize their outcomes, and y_i is the outcome of agent *i*)[69].

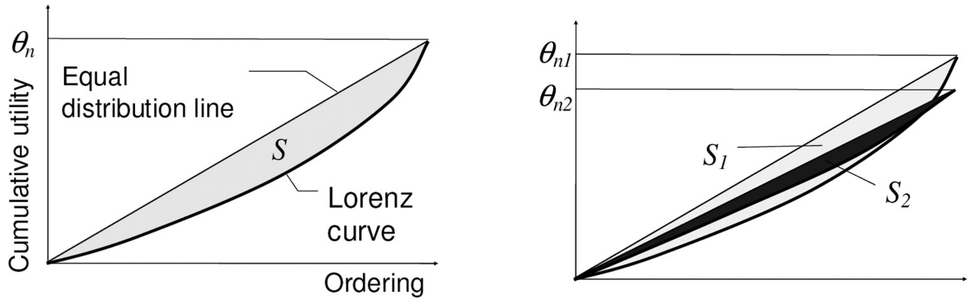
Note that this formulation of the distribution problem does not use subjective agent utilities, but rather uses objective criteria that are the same for all agents (for example, if the problem is a distribution of goods, than an objective criterion could be the monetary value of the goods at a market price; although it is possible that agents would subjectively value some goods higher in spite of a lower or equal monetary value). However, the theory of equity can also be formulated using subjective utilities, under the assumption that these utilities are comparable (Lissowski 2008; Sen 1970). The following explanation of the theory of equity applies in both cases, but we have chosen to present it using objective criteria because of increased simplicity.

Note that we assume that all agents are equally entitled or capable of achieving good outcomes. We shall call such agents *similar agents*. The theory of equity can be extended to take into account various priorities of agents, but this makes the definition considerably more complex (Ogryczak 2009). If the agents are not similar because they have different levels of expenditure or contribution and are therefore entitled to different outcomes, a common practice is to transform every agent's outcome by dividing them by the agent's contribution (Wierzbicki, Kaszuba et al. 2009). After such a transformation, it is possible to think of the agents as similar, because they are equally entitled to receive a unit of outcome per unit of contribution. If some agents are not similar for other reasons (in an Internet auction, the reason can be that various sellers have various quality of goods or services, and various marketing), then it is still possible to consider the fairness for a subset of agents that are similar according to these criteria. A system should be able to at least provide fairness to this subset of similar agents. This approach is equivalent to a ceteris paribus assumption from economics: when *all other factors can be excluded* and all agents are equally entitled, the theory of equity can be used for testing distributional fairness. In a laboratory setting, such conditions can be satisfied and we can design systems that realize the goal of fairness, even in the presence of adversaries that do not act in a procedurally fair manner.

Figure 3.1 shows one of the key concepts illustrating distributive fairness: the Generalized Lorenz curve. The Generalized Lorenz curve is obtained by taking the outcomes of all agents that participate in a distribution and ordering them from

[69] An optimal solution of this problem is any Pareto-optimal solution: a solution with the property that it is not possible to improve any of its outcome values without worsening another.

worst to best (the Generalized Lorenz curve is usually divided by the number of agents, n [Shorrocks 1983]. In this text, Iuse a rescaled version that is not divided by n). Let us denote this operation by a vector function $\theta(y) = [\theta_1(y),..., \theta_n(y)]$ of the outcome vector y.



a) The Lorenz Curve

b) Two incomparable distributions

Fig. 3.1. Examples of Lorenz curves

Then, the cumulative sums of agents' utilities are calculated: starting from the utility of the worst agent (θ_1), then the sum of utilities of the worst and the second worst (θ_2), and so on, until the sum of all agents' utilities. This sum is denoted on the figure as θ_n. The second line on the figure, the equal distribution line, is simply a straight line connecting the points $(1, \theta_1)$ and (n, θ_n). The area between the two curves, denoted by S, can be seen as a measure of inequality of the agent's utilities. The objective of distributive fairness is to reduce this inequality, bringing the Lorenz curve closer to the equal distribution line, while at the same time keeping in mind the total efficiency (sum of all agents' utilities), which is represented by the right end of the Lorenz curve at a value of θ_n. Note that these two objectives frequently form a trade-off (for example, if the distribution is constrained by a budget).

The right part of the figure 3.1 shows two Lorenz curves that correspond to different distributions among the same agents. The first distribution has a higher θ_n, but also a higher inequality, while the second distribution has a lower total of agents' utilities, but is more fair. In terms of equitable optimization, the two distributions on the right part of the figure 3.1 are incomparable – the choice of one of them depends on the preferences of a decision maker[70]. However, the most desirable goal of the theory of equity is finding solutions that are *equitably optimal*. For any solution that is not equitably optimal, we can find another solution such that its Lorenz curve is at every point above the Lorenz curve of the equitably dominated solution.

The area between the Lorenz curve and the equal distribution line can be simply calculated and used as a computable measure of inequality. The Gini coefficient (frequently used in economics) is the area S normalized by θ_n: $\mathrm{Gini} = S/2\theta_n$. Note that minimizing the Gini coefficient can lead to worse total outcomes (sums of all agent's

[70] Incomparable distribution can also have identical total efficiencies θ_n.

utilities). This drawback can be overcome by taking into account the Gini coefficient and the sum of all utilities at the same time. When two distributions are compared, if one of them has a smaller Gini coefficient and a larger sum of all utilities, then it should be more equitable (although this is not a sufficient condition).

It is also possible to use a different inequality measure: *the area below the Lorenz curve* (equal to $BLC = (n\theta_n/2) - S$). In this chapter, I have chosen to use the BLC as an inequality measure. The reason for this choice is that an improvement in terms of the theory of equity (a new, equitably dominating solution) always causes an increase of the BLC. On the other hand, an equitably dominating solution can have a larger Gini than a dominated solution (consider for example a solution that just increases the outcome of the best-off agent. This solution may dominate another, but will have a larger Gini and a larger BLC). Therefore, the area below the Lorenz curve is the best criterion of inequality, according to the theory of equity. However, note that even if the area below the Lorenz curve increases, the distribution does not always become more equitable. An increase of the BLC and the total sum of outcomes gives a better assurance that the distribution is indeed more equitable, although it is still not a sufficient condition (only by considering all partial sums that form the generalized Lorenz curve are we able to verify equitable domination with certainty).

It is necessary to use all fairness criteria with caution, since finding *equitably optimal solutions* is a multi-criteria problem that cannot be simply reduced to the comparison of single or two criteria. According to the theory of equity, an *equitably optimal solution* (or simply equitable solution) is any Pareto-optimal solution of the *equitable optimization problem*, which is obtained from the efficient optimization problem by the transformation θ of cumulative ordered sums. The criteria of the equitable optimization problem are simply the vector θ_y. The theory of equity allows not only to define fairness with precision, but also to search for equitable solutions by applying standard methods of multi-criteria optimization to the equitable optimization problem.

The theory of equity also has an axiomatic expression (Kostreva, Ogryczak and Wierzbicki 2004). The axioms of the theory of equity define a preference relation on the outcome vectors of the efficient optimization problem. An *equitable preference relation* is any symmetric and transitive relation satisfying the following axioms (Kostreva, Ogryczak and Wierzbicki 2004):

Symmetry – The ordering of the outcome values is ignored (e.g. a solution $y = [4, 2, 0]$ is equally good as a solution $y = [0, 2, 4]$). First of all, fairness requires impartiality of evaluation, thus focusing on the distribution of outcome values while ignoring their ordering. That means, in the efficient optimization problem we are interested in a set of outcome values without taking into account which outcome is taking a specific value. Hence, we assume that the preference model is impartial (anonymous, symmetric). In terms of the preference relation it may be written as the following axiom for any permutation τ of I:

$$(y_{\tau(1)}, y_{\tau(2)}, \ldots, y_{\tau(n)}) \cong (y_1, y_2, \ldots, y_n), \tag{9}$$

which means that any permuted outcome vector is indifferent in terms of the preference relation.

Monotony – A outcome improving the value of one of the objectives is preferred, the values of other objectives are not deteriorated (e.g. y = [4, 2, 0] is preferred to y = [3, 2, 0]). This axiom is actually a repetition of the requirement of efficiency. It guarantees that only efficient solutions will be chosen as equitable solutions. Another way of looking at it is that the monotony axiom prevents a phenomenon well-known in former Communist countries: that of "equating downwards", or making outcomes worse (and more equal, but not more equitable) for everyone. The same phenomenon could also occur if agents can cheat on effort, such as freeriders in P2P systems. The monotony axiom assures that a system that is dominated by freeriders will not be considered as good as a system where some peers provide content. The axiom can be expressed as follows:

$$\mathrm{y} - \epsilon e_i \succ y \quad \text{for} \quad \epsilon > 0, \;\; 1 \leq i \leq n \tag{10}$$

Principle of transfers – A transfer of any small amount from an outcome to any other relatively worse-off outcome results in a more preferred outcome vector (e.g. y = [3, 2, 1] is preferred to y = [4, 2, 0]). Fairness requires equitability of outcomes which causes that the preference model should satisfy the (Pigou–Dalton) principle of transfers. The principle of transfers states that a transfer of any small amount from an outcome to any other relatively worse-off outcome results in a more preferred outcome vector. As a property of the preference relation it represents the following axiom

$$y_{i'} > y_{i''} \;\Rightarrow\; \mathrm{y} - \epsilon e_{i'} + \epsilon e_{i''} \succ y \quad \text{for} \quad 0 < \epsilon < y_{i'} - y_{i''} \tag{11}$$

It can be shown that any *equitably optimal solution (a Pareto-optimal solution of the problem θ_y) is not dominated by any other solution in the equitable preference relation* (Kostreva, Ogryczak and Wierzbicki 2004). Thus, the concept of distributive fairness as expressed by the theory of equity is well understood by considering the three above axioms. These axioms show, among other things, that the theory of equity avoids the pitfall of preferring more equal, but less globally efficient solutions. According to the axiom of monotony, increasing any objective without worsening the others improves the overall solution in terms of the equitable preference relation.

Using the theory of equity, the Fairness Emergence hypothesis can be reformulated as follows: *if a good trust management system is used by agents, then distribution of similar agents' utilities should become more equitable*. The inequality criteria described in this section, such as the Gini coefficient or the area below the Lorenz curve (BLC), can be used together with the total efficiency θ_n to check whether a distribution has become more equitable (in rare cases, these two criteria may not be sufficient to guarantee equitable domination, but that can be done only by checking all criteria of the equitable optimization problem). Taking into account the total efficiency alongside with BLC or Gini will allow us to check whether the reputation system is capable of finding equitable solutions that are also good in the utilitarian sense.

3.1.3. Design of Simulation Experiments

To verify the Fairness Emergence hypothesis, we have used simulation experiments. The FE hypothesis would hold if we could establish that the reputation system causes an increase of the equity of the distribution of utilities. In particular, we will be interested to study the impact of the quality of the reputation system on the equity of utility distributions.

The simulator is based on the Repast 3.1 platform[71] and resembles an Internet auction system. In the design of the simulator, we had to make a decision about a sufficiently realistic, yet not too complex model of the auction system, of user behaviour, and of the reputation system. We chose to simulate the reputation system and the behaviour of its users as faithfully as possible (the only simplification is that we use only positive and negative feedbacks).

The auction system, on the other hand, has been simplified. We simulate the selection of users using random choice of a set of potential sellers. The choosing user (the buyer) selects one of the sellers that has the highest reputation in the set.

After the buyer has selected a seller, a transaction between the two agents may occur. However, this is not always the case in our simulations, because the chosen seller may have a reputation that is too low for the buyer. If the chosen seller has a reputation below the buyer's acceptance threshold, no transaction will occur. Still, we count the number of such *transaction attempts* in our simulation. The number of transaction attempts is used as a measure of time in our simulation (since we assume that each transaction attempt would consume some time and effort on behalf of a buyer, the number of such transaction attempts is limited). Furthermore, the granularity of transaction attempts in our simulation is very high. To show meaningful results, we group several hundred subsequent transaction attempts into one *turn*. The turn is used as a measure of time for the demonstration of simulation results.

In this chapter, I describe two sets of simulation results. The first set was obtained from simulations of a closed system of agents – the set of agents was kept fixed for the duration of the simulation. This approach has been used initially to reduce the number of factors that could impact the results, and to study fairness emergence in a simpler setting. The second set of simulation results was obtained from a trace-driven simulation approach that was used to control the presence of seller sin the system. This allowed for a more realistic simulation of an open system of buyers and sellers, where the time a buyer or seller could spend in the system was controlled

[71] http://repast.sourceforge.net/repast_3/index.html (1.03.2012).

by the trace. The second set of simulation results takes into account more complex factors, but was used to verify the results from the first set.

3.1.3.1. Agent behavior

In our simulator, a number of agents interact with each other. There are two types of agents in the system: fair and unfair agents. Dishonest agents model adversaries. To test the FE hypothesis, we shall be interested in the fairness of utility distributions of fair agents. The payoffs of fair and unfair agents will also be compared.

In the closed system simulations, all agents are similar. In the trace-driven simulations that will be discussed in more detail below, agents can be buyers or sellers (this separation is a consequence of the separation of roles in real auction systems, where users mostly either buy or sell). This additional distinction makes the simulations more realistic.

When an agent wants to carry out a transaction, it must make three decisions. The first decision concerns the choice of a transaction partner (seller) and whether or not to engage in the transaction. The agent chooses his partner from a randomly selected set of k other agents (in the simulations of the closed system, k has been equal to 3 or 1). From this set, the agent with the highest reputation is chosen. However, if the highest reputation is lower than a threshold $p_{min}choice$ (in the closed system simulations, fair agents choose partners with reputation at least 0.45, and unfair agents: 0.3),then the choosing agent will not engage in any transaction. If the best agent's reputation is sufficiently high, the choosing agent will engage in the transaction with a certain probability p (in the simulations presented here, this probability was 1).

The second decision concerns the agent's behavior in the transaction. This decision can be based on a game strategy that can take into consideration the agent's own reputation as well as the reputation of his partner, the transaction history and other information. We decided to use the famous Tit-for-tat strategy developed by Rapaport but extended with using a reputation threshold: if two agents meet for the first time and the second agents' reputation is below $p_{min}game$, the first agent defects. The strategy used in the simulations presented here has also been based on the threshold $p_{min}cheat$. In the case when the partner's reputation is higher than $p_{min}cheat$, the agent would act fairly; otherwise, it would cheat with a certain probability c. In the simulations presented here, fair agents had a cheating probability of 0, while unfair agents had a cheating probability of 0.2 and a reputation threshold of 0 – meaning that unfair agents cheated randomly with a probability of 0.2.

The third decision of the agent concerns the sending of reports. For positive and negative reports, an agent has separate probabilities of sending the report. In the simulations presented here, the probability of sending a positive report, p_{rep+} was 1.0, while the probability of sending a negative report p_{rep-} varied from 0 to 1.This choice is based on the fact that in commonly used reputation systems (Wierzbicki

2006), the frequency of positive reports is usually much higher than of negative reports. In the simulation it is also possible to specify a number of agents that never send reports. This behaviour is independent of the honesty or dishonesty of agents.

The strategies of agents in our simulations do not evolve, but remain fixed for the duration of simulation. In this respect our work is different from the research on evolution of cooperation or indirect reciprocity (Wilson 1985). Our research is focused on verifying the effect of trust management on fairness, without considering how the strategy of using trust management or reputation has evolved – that is the concern of related and future work (Pollock 1992; Castelfranchi 1988).

Note here that the presented model of agent behaviour with respect to the reputation system matches many kinds of applications. The model has been described using Internet auctions as an example. Another kind of realistic application is a Peer-to-Peer system. A transaction in such a system is an exchange of data or services (resources). Unfair behaviour in such a system is called free-riding: peers use resources of others, but do not reciprocate. A P2P application can use reputation to combat free-riding. The reputation system in a P2P application is distributed, in contrast to the reputation system used in Internet auctions. However, the discovery of proofs by the P2P reputation system is affected by the quality of the distributed search algorithms and by the presence of adversaries, who can attempt to drop negative proofs. A result is a smaller availability of negative reports, which has been modeled in the simulator by varying the probability p_{rep}- varied from 0 to 1. This type of adversary has been discussed frequently in the literature (Liu 2004; Lee, Sherwood and Bhattacharjee 2003; Kamvar 2003). The first set of our simulations presents results that can apply also to P2P applications that use a reputation system.

3.1.3.2. Reputation system warm-up

A real reputation system has a large initial history that can be used to evaluate infrequently present agents. In the simulation approach, this initial history had to be reproduced. In the closed system, for each simulation, the first 20 turns have been used to warm-up the reputation system by acquiring an initial history of agent behaviour. This means that the payoffs have not been recorded, but an agents' reputation has been modified by positive and negative reports. This method has been used to model the behaviour of a real reputation system, where the system has available a long history of transactions. Simulating the reputation system without a warm-up stage would therefore be unrealistic.

In a closed system, it is possible to warm-up the reputation of all agents at the same time, at the beginning of the simulation. In the open, trace-driven approach, the trace represents a period of time taken from the operation of a real Internet auction site. Agents present in the trace could have been present in the system before the beginning of the trace. As this information is not available, it is also not realistic to simulate the system without a warm-up. However, this warm-up can be done separately for each seller. If a buyer would select a seller that was in the warm-up stage, the results of the transaction were not recorded in the utility of the buyer and the seller. The seller's reputation was updated. A fixed number of l transactions was used as a warm-up. This ensured that the reputation system had some initial information about each seller, before the buyers utilities were recorded. In the simulation results presented below, $l = 5$. Reducing the length of the warm-up had a strong effect on emergence: emergence was not observed for $l = 0$, for any other setting of simulation parameters.

3.1.3.3. Trace-driven simulation of an Internet auction system

Trace-driven simulation allows to overcome two main drawbacks of simpler simulation approaches. The utilities of agents in the system will depend on the time that agents spend in the system. The simplest approach would be to simulate a closed system of agents; however, such an approach may not be sufficiently realistic, as agents in real ODS tend to join and leave the system frequently. This limitation may be removed by allowing agents to be in the system for a random number of rounds. This time of an agents' activity can chosen from a distribution that is similar to empirical data (for example, a Pareto distribution). Yet, this method of simulating an agent's activity is still not sufficiently realistic. For that reason, we have decided to use trace-driven simulations.

We have obtained a trace from the largest Polish Internet auction site. The trace includes approximately 200 000 seller transactions from 6 months. In the trace, there were about 10000 sellers randomly selected from the auction house. The weekly number of seller transactions in the trace is shown on the figure 3.2. The trace was used to control the times spent by sellers in the simulated system. In other words, using trace-driven simulations allowed us to simulate an open system and to preserve the real processes of seller activity in the system. Figure 3.3 shows the distribution of the number of transactions made by a seller. It can be seen that this distribution resembles a heavy-tailed distribution.

The behavior of sellers was not recorded in the trace, and it is therefore simulated as described in the previous section. Moreover, the buyers are not trace-driven. Buyers initiate transactions with sellers who are present in the system at a given time (in the trace-driven simulations, one turn is equivalent to one day of the trace. During this turn, only the buyers who offered auctions on that day are present in the system). Buyers choose sellers in the same way as in the simulations of the closed system, choosing a seller with the highest reputation from a random set of k sellers that are active in this turn. Buyers are also able to reject transactions if a chosen seller's reputation is below a threshold.

The second drawback of simple simulation approaches is related to the lack of roles of agents. In the trace driven simulation, it was possible to divide agents into two sets of buyers and sellers, which allowed the simulations to resemble real

Internet auctions. The proportions and activity distributions of buyers and sellers were preserved.

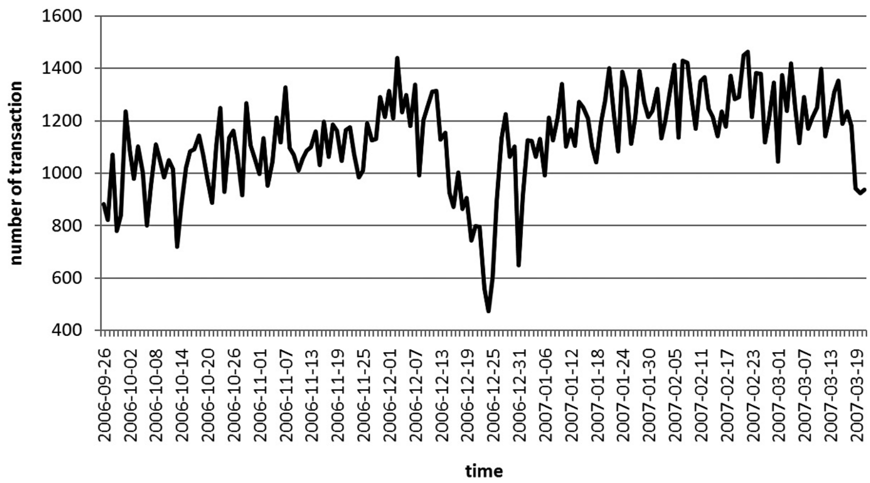


Fig. 3.2. Daily number of seller transactions in the trace

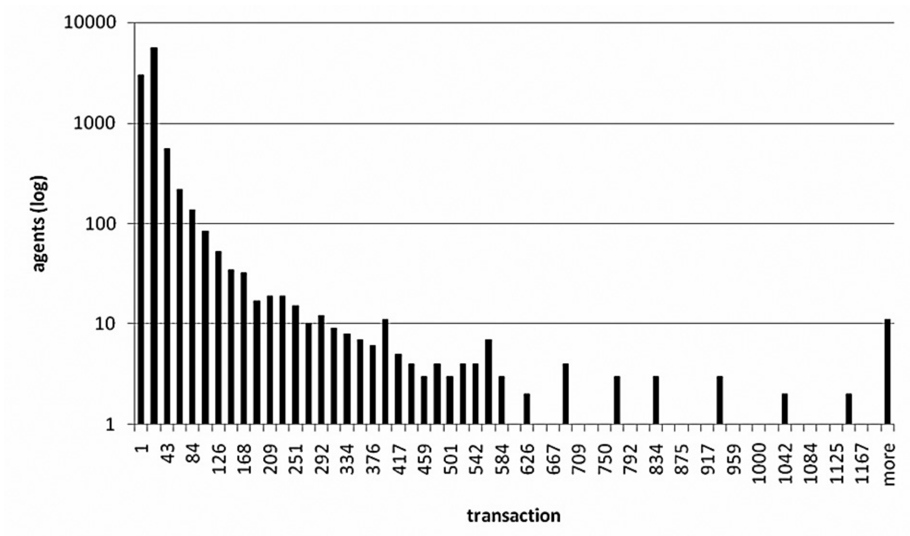


Fig. 3.3. Distribution of number of transactions of a seller

3.1.4. Fairness emergence in a closed system

3.1.4.1. Experiment setup

In simulations of the closed system, there was a total of 1500 agents, out of which 1050 where fair and 450 were unfair. While the proportion of unfair agents is high, they cheat randomly andat a low probability – so an unfair agent is really a "not totally fair agent". Also, considering that frauds in Internet auctions are among the most frequent digital crimes today, and considering that cheating in a transaction may be more frequent than outright fraud – it may be sending goods that are of worse quality than advertised – this proportion of unfair agents seems realistic.

The simulator can compute reputations using all available feedbacks. The results of the simulation include: the reputations of individual agents and the total utilities (payoffs from all transactions) of every agent. In the simulations presented here, an agent's reputation is computed as the proportion of the number of positive reports about the agent to the number of all reports.

All simulations were made using pseudo-random numbers, therefore the Monte Carlo method is used to validate statistical significance. For each setting of the simulation parameters, 50 repeated runs were made, and the presented results are the averages and 95% confidence intervals for every calculated criterion. The confidence intervals were calculated using the t-Student distribution.

We decided to use transaction attempts instead of the number of successful transaction as a stop condition because we believe that an agent would consider each transaction attempt as an expense of time and effort. In most presented simulations for each turn, 500 transaction attempt shave been made.

3.1.4.2. Closed System Simulation Results

To verify the Fairness Emergence hypothesis, we have been interested to investigate the impact of a reputation system on the equity of the agent utility distribution. Equity of utility distributions has been measured using fairness criteria based on the theory of equity; however, other criteria such as the sum of agent utilities are considered as well. The simulations revealed that the Fairness Emergence hypothesis holds in several cases, but not universally; therefore, we have investigated the sensitivity of fairness emergence to various factors that influence the quality of the reputation system.

Fairness Emergence in the Long Term

The first studied effect has been the emergence of fairness in the long term. In the simulation experiment, we have measured the area under the Lorenz curve (BLC) and have run the simulation until the BLC stabilized. This experiment has been repeated using three scenarios: in the first one, the agents did not use any reputation system, but selected partners for transactions randomly. In the second experiment, the reputation system was used, but agents submitted negative reports with the probability of 0.2. In the third experiment, negative reports have always been submitted.

The results of the three experiments are shown on the figure 3.4. The figure plots the average BLC of fair agents from 50 simulation runs against the number of turns of the simulation. It can be seen that when agents do not use the reputation system, the BLC stabilizes for a value that is almost twice smaller than the value of BLC that is obtained when reputation is used. Furthermore, there is a clear effect of increasing the frequency of negative feedbacks: the BLC increases faster and stabilizes at a higher value when $p_{rep-} = 1$. The initial decrease of the BLC from 1 is due to the fact that at the beginning of the simulation, the distribution of fair agent utilities is equal (During the warm-up stage, utilities of agents are not recorded. All agents start with a zero utility after warm-up completes).

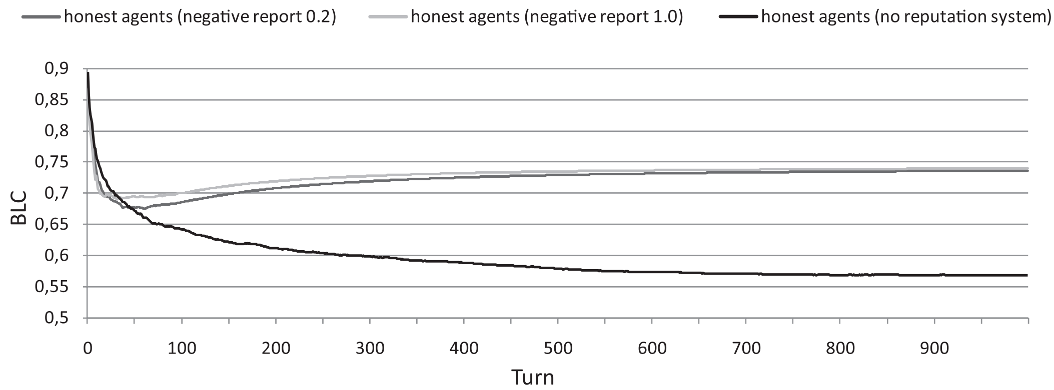


Fig. 3.4. Fairness Emergence in the long term

The result of this experiment seems to be a confirmation of the FE hypothesis. The distributions of fair agents' utilities have a higher BLC (and a higher total sum of utilities) when the reputation system is used. Yet, the problem here is that in realistic auction systems, most agents only have a small number of successful transactions, because they use the system infrequently. In our simulation, new agents did not join the system (although the number of agents was large). The average number of successful transactions of an agent has been about 270, which is much lower than the number of agents; this means that as in a real auction system, the chance of repeated encounters was low. However, this number is still large. The simulations were continued until a stable state was reached; in practical reputation systems, such a situation would not be likely to occur because of the influx of new agents and the inactivity of old ones. For that reason, we have decided to investigate the FE hypothesis in the short term, or in unstable system states.

Fairness Emergence in the Short Term

The simulation experiments used to study short-term system behavior have been about 8 times shorter than the long-term experiments. For these experiments, the number of successful transactions of an average agent was about 60. Figure 3.5 shows the BLC of the distributions of fair agents' utilities. On the x axis, we show the number of turns. The figure shows two lines corresponding to different frequencies of sending negative reports by fair agents (unfair agents always sent negative reports). The results show that for low negative report frequencies fairness emerges more slowly. Increasing the available negative reports reduces the time needed for fairness emergence. This effect is apparent very quickly, even after 50 turns of simulation.

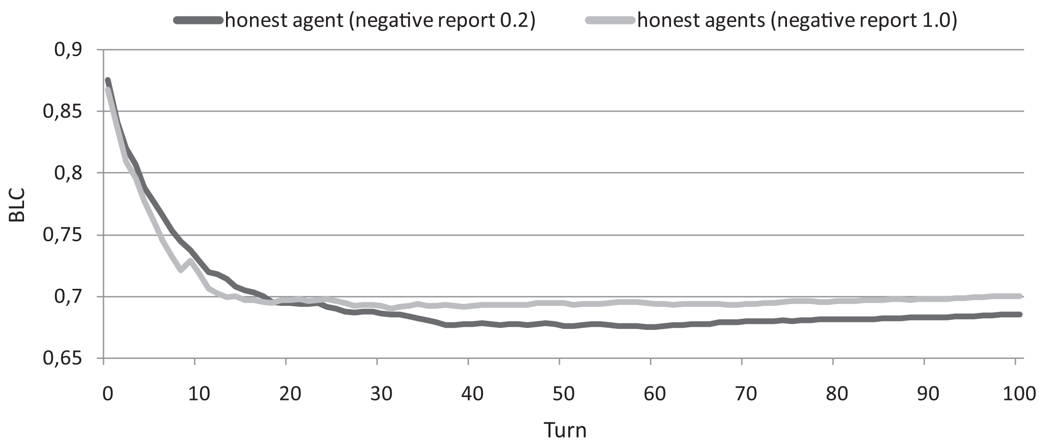


Fig. 3.5. Fairness Emergence in the short term

From now on, fairness emergence in the short term is studied more closely to verify whether the improvement of reputation system quality will strengthen fairness emergence. In other words, until now we considered fairness emergence with time, and now we shall consider the *sensitivity of fairness emergence to the reputation system's quality*. All further experiments have been made in the short term, outside of the stable state of the system.

3.1.4.3. Effect of Better Usage of Reputation

The usage of reputation by agents had a particularly strong influence on the emergence of fairness. In our simulations, during a transaction attempt, agents chose a seller with the highest reputation from a set of k candidates. The chosen candidate needed to have a reputation that was higher than the buyer's threshold. If $k = 1$, then the transaction partner was chosen at random and only the threshold $p_{min}game$ was used to consider reputation. If $k = 3$, it was less likely that an agent with lower reputation would be chosen as a transaction partner. These two scenarios correspond to the real life situation of buyers who are able to select sellers from a larger set, based on their reputation; on the other hand, it could be possible that the choice is low, because only one seller has the required goods or services.

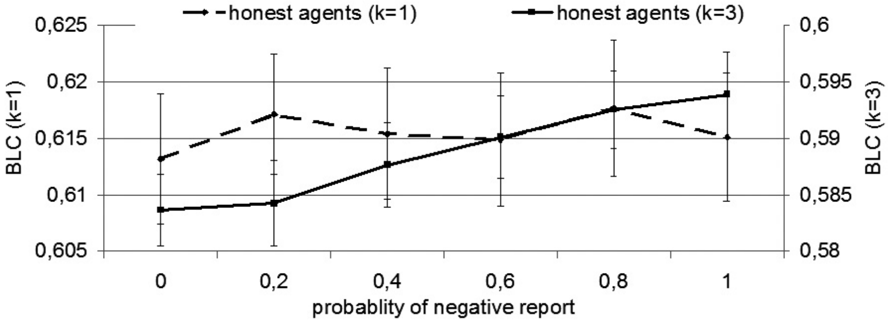


Fig. 3.6. Effect of increased choice on BLC

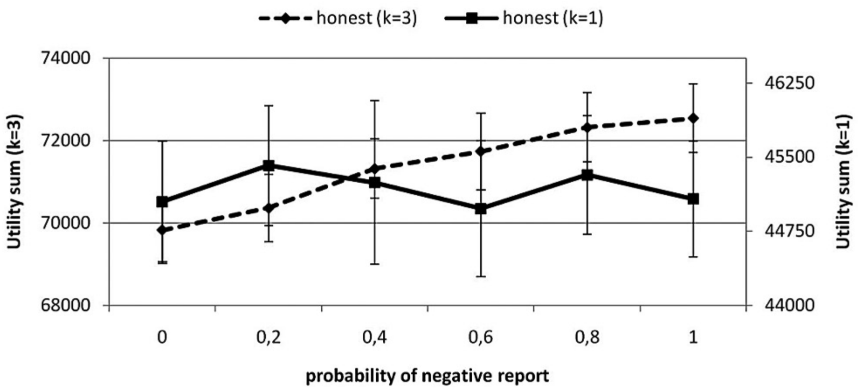


Fig. 3.7. Effect of increased choice on sum of utilities

We have considered the two scenarios while investigating the impact of the frequency of feedbacks on the reputation system. It turns out that increasing the choice of agents is necessary for the emergence of fairness. Figure 3.6 shows the effect of increasing the frequency of negative feedback on the BLC of fair agents. The figure shows two lines that correspond to the scenarios of $k = 1$ and $k = 3$. It can be seen that if the choice of agents on the basis of reputation is possible ($k = 3$),then the increase in the number of feedbacks leads to a increase of BLC. On the other hand, if the choice is limited ($k = 1$), then the increase in the number of negative feedbacks does not have a statistically significant effect on the BLC. This effect is best explained by the fact that if choice is available, honest agents have a better chance at avoiding dishonest agents while at the same time they do not waste transaction attempts. If agents do not have choice, they can still avoid transactions with dishonest agents, but they will waste transaction attempts and have a lower utility.

Figure 3.7 shows the effect of increased choice and varying negative feedback frequency on the sum of fair agents' utilities. It can be seen that once again, enabling the choice of partners based on reputation has a positive effect on the welfare of fair agents. For $k = 3$, fair agents overall had a higher sum of utilities than for $k = 1$, and this sum increased when the frequency of negative reports increased. This also explains why the BLC of fair agents for $k = 1$ was higher than for $k = 3$. Since the sum of utilities was lower for $k = 1$, the BLC could also be lower, although this does not mean that the distribution of utilities for $k = 1$ was more equitable than for $k = 3$.

3.1.4.4. Effect of Better Feedback

Better feedback is a prerequisite for increasing the quality of a reputation system. For that reason, we have chosen to investigate the effect of increased feedback on the emergence of fairness. As has been explained previously, the frequency of negative feedback has been varied from 0 to 1. We have also varied the frequency of positive feedbacks and negative feedbacks simultaneously; however, for the simple reputation algorithms considered in this chapter, the only significant parameter is the proportion of negative to all feedbacks. For that reason, varying negative feedbacks' sending frequency is sufficient to evaluate the system's sensitivity to feedback availability. Another issue related to feedbacks is the possibility that agents send false feedbacks. Our studies indicate that a small amount of false feedbacks does not impact the results, but a significant amount of false feedbacks will confuse any reputation system. For this reason, in this analysis we disregard the possibility of sending false feedbacks by the agents.

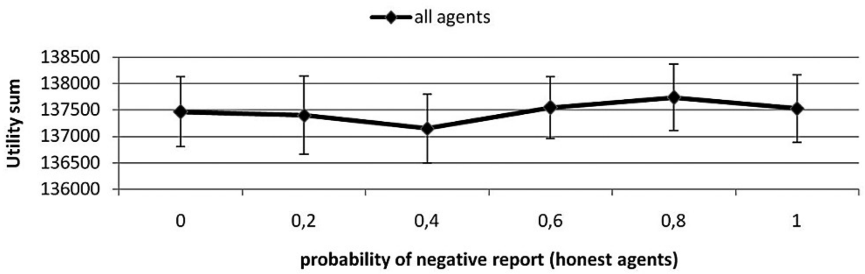


Fig. 3.8. Effect of increased feedback on sum of utilities of all agents

Figure 3.8 shows the effect of increasing the frequency of sending of correct negative feedback on the sum of utilities of all agents. If the frequency of sending negative feedback is 1, then the reputation system receives all information relevant to the computation of correct reputations. If the frequency is low, then the reputation system is missing important information that could decrease the reputations of unfair agents.

It turns out that the total sum of all agents' utilities was not affected by the increase of negative feedback frequency. This seems to be a paradox, since we are using the iterated Prisoner's Dilemma as a model of our auction system. Increasing negative feedbacks from 0 to 1 should result in decreasing the ability of unfair agents to cheat, and should therefore increase the payoffs received by both agents in a transaction. However, the number of transactions may be affected, as well, and this can explain the apparent paradox.

This experiment also shows that even assuming the use of a Prisoner's Dilemma as a model of a transaction, the use of the sum of all agents' utilities (the utilitarian paradigm) would lead to a wrong conclusion that the system behaviour is not affected. From the utilitarian point of view, the reputation system works equally well when the frequency of negative reports is 0, as when it is equal to 1.

Figure 3.9 shows that this is not the case. The sum of utilities of fair agents increases, as negative feedbacks are sent more frequently. On the other hand, the sum of utilities of unfair agents drops. The reason for this fact is that with higher frequencies of negative feedback, the reputations of unfair agents decrease, and therefore these agents have fewer successful transactions. On the other hand, fair agents manage to avoid unfair ones, and do not waste transaction attempts (therefore they have more successful transactions and higher payoffs in these transactions).

Figure 3.10 shows effect of increased negative feedback frequency on the BLC. Clearly, increased negative feedback frequency leads to an increased BLC of honest agents' utilities. Note that the effect is statistically significant for the variation of from 0 to 1 (also from 0.4 to 1). Note that these simulations have been made in the short term and that together with the results about the sum of utilities, they prove the FE hypothesis: increasing the quality of the reputation system does indeed lead to more equitable distribution of fair agents' utilities, as the hypothesis suggested.

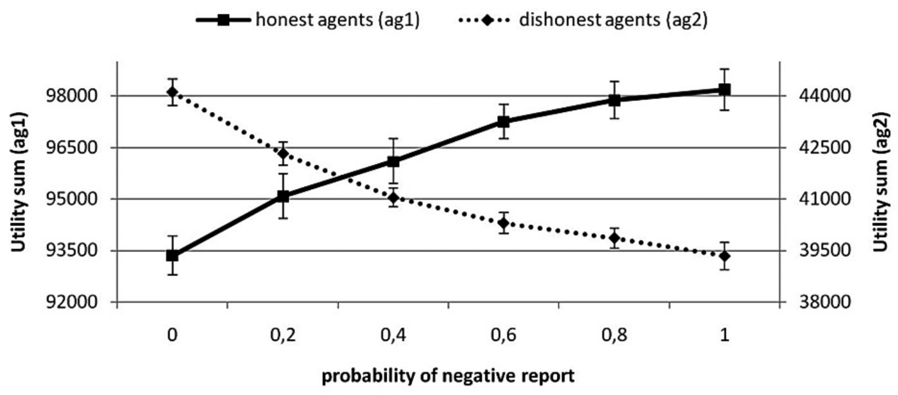


Fig. 3.9. Effect of increased feedback on fair and unfair agents' utilities

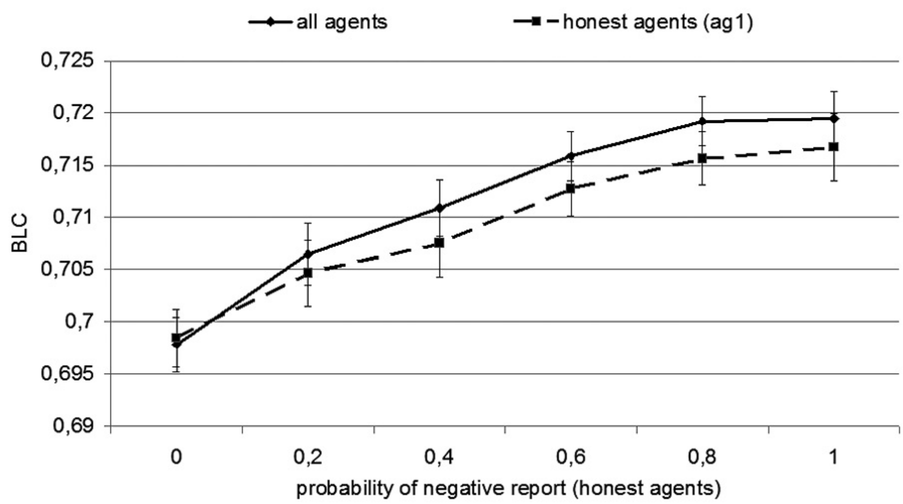


Fig. 3.10. Effects of increased feedback on BLC

3.1.5. Effect of Improved Reputation Algorithm

Fairness emergence could be sensitive to a change in the algorithm that is used to calculate reputations. With better algorithms, perhaps it would be possible to improve fairness. That would be equivalent to fairness emergence with improved trust management system's operation.

3.1.5.1. Algorithm of implicit negative feedbacks

The algorithm described in this section has been introduced (Wierzbicki 2006). Most online auction sites use a simple feedback-based reputation system (Resnick, Zeckhauser 2002). Typically, parties involved in a transaction mutually post feedbacks after the transaction is committed. Each transaction can be judged as "*positive*", "*neutral*", or "*negative*". The reputation of a user is simply the number of distinct partners providing positive feedbacks minus the number of distinct partners providing negative feedbacks (possibly normalized by the number of all distinct partners). As pointed out in (Malaga 2001), such a simple reputation system suffers from numerous deficiencies, including the subjective nature of feedbacks and the lack of transactional and social contexts. Yet another drawback of feedback-based reputation systems is that these systems do not account for psychological motivation of users. Many users refrain from posting a neutral or negative feedback in fear of retaliation, thus biasing the system into assigning overestimated reputation scores. This phenomenon is manifested by high asymmetry in feedbacks collected after auctions and, equally importantly, by high number of auctions with no feedback provided. Many of these missing feedbacks may convey implicit and unvoiced assessments of poor seller's performance which should be included in the computation of a seller's reputation.

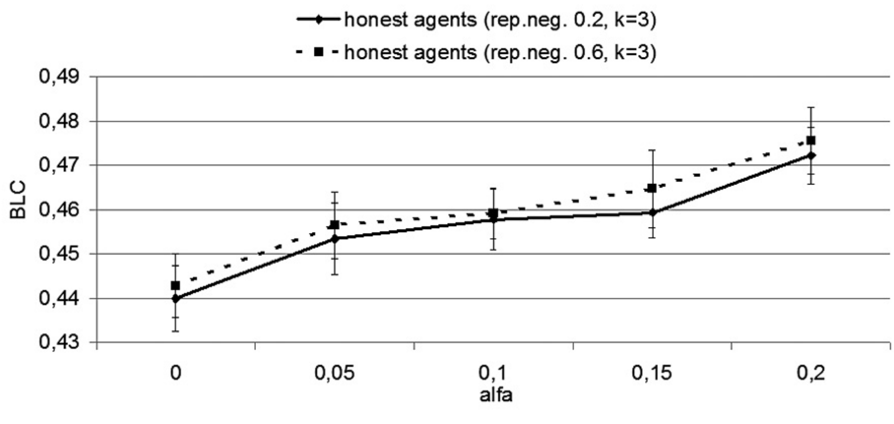


Fig. 3.11. Effect of improving reputation algorithm on BLC

As described in (Wierzbicki 2006), there can be many ways of identifying implicit feedbacks in a real-world reputation system, based on the observation of behavioural patterns. To evaluate the effectiveness of using implicit feedback, we have identified a simpler reputation algorithm that can be simulated and compared to the algorithm of most Internet auction houses.

Consider a user u with a history of n auctions. Let us assume that only $m \leq n$ of these auctions have a feedback. Out of these m feedbacks m^+ are positive or neutral feedbacks (in practice, neutral feedbacks in reputation systems have a meaning almost similar to negative feedbacks and are very rare), m^- are negative feedbacks, while $m^* = n - m$ is the amount of missing feedbacks (transactions that had no feedback). Thus, $m^+ \leq m \leq n$. The reputation ρ_u of the user u will be calculated as follows:

$$\rho_u = \frac{m^+}{\alpha m^* + m^+ + m^-}, \text{ where } 0 \leq \alpha \leq 1 \tag{12}$$

Thus, if $\alpha = 0$, the above reputation score becomes a simple ratio of the number of positive feedbacks received by the user u. In the case when the user has had no auctions, the above formula is undefined. In such case we set the reputation ρ_u to an initial value, ρ_0. The coefficient α is used to control the importance of implicit negative feedbacks.

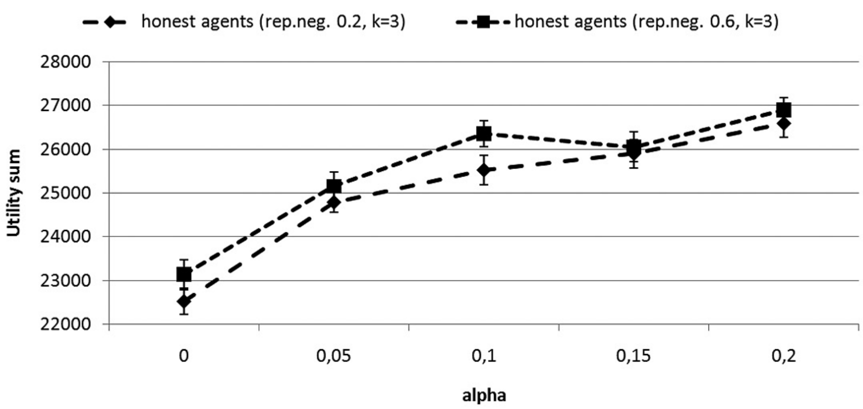


Fig. 3.12. Effect of improving reputation algorithm on utility sum

To be precise, in our simulations we use a slightly more complex version of the above algorithm. Since agents in the simulator choose whom they want to interact with on the basis of reputation scores, it is necessary to avoid that the reputation would drop suddenly to a low level. This can happen in the initial phase of the simulation, when the reputation score has not yet stabilized (initially, a single negative feedback could decrease the initial reputation by a large degree). Therefore, we use a simple moving average to smooth reputation changes. The smoothed reputation $\rho_u^{ma}(t) = 0.5\rho_u^{ma}(t-1) + \rho_u(t)$ where t is time, and $\rho_u^{ma}(0) = \rho_0$ (the smoothed reputation is initialized by the initial reputation value). Note that over time, the impact of the initial reputation decreases exponentially.

The results of increasing α from 0 to 0.2 on the BLC of honest agents' utility distribution and on the sum of honest agents' utilities are shown on figure 3.11 and figure 3.12, respectively. The figures show several lines that correspond to various frequencies of negative reports. Increasing the role of implicit negative feedbacks clearly increases fairness, and the effect is strong and statistically significant. This behaviour is a confirmation of the FE hypothesis in an unstable state, in the presence of adversaries, and when the probability of negative reports is low. The sum of honest agents' utilities also increases when α is increased.

The observed effect can be explained similarly as the effect of increasing frequencies of sending negative feedback. Note that the effect of varying α from 0 to 0.2 is similar to the effect of increasing the probability of negative feedbacks. Larger values of α have been found to lead to an decrease of the Gini coefficient in our previous research (Wierzbicki 2007) but this effect has been obtained for a different, specific version of the simulations system and need to be studied further to allow generalization.

3.1.6. Fairness emergence in the open system

In the previously described simulations scenarios, all agents were treated equally (although we have referred to the agent who initiated a transaction attempt as a "buyer", all agents had an equal chance to become a "buyer" in the described simulations). Moreover, all agents had a similar level of activity in the system. Agents could not leave the system during the simulation and were chosen over and over again for transaction attempts. New agents could not join the system. This approach had an impact on the evaluation of the reputation system. In a realistic reputation system, the amount of information available about new agents would be considerably less than the amount of information available about agents that have been active for some time in the system. In the closed system, in the long term, the reputation system would have very good information about agents. Considering the operation of the reputation system in the short term partially reduces that problem, but does not fully solve it.

The reason for the use of the closed system is that without additional information or assumptions, it was not possible to specify how active should the agents be in the system. In this section, we are going to remove this limitation, based on the trace-driven approach described in the previous section. On the other hand, the previously described results were more general and could apply to a variety of applications of reputation systems.

To test the FE hypothesis in an open system, we have measured the utilities of two kinds of agents: the buyers and the fair sellers. By the FE hypothesis, the distribution of both kinds of utilities should become more equitable with an improvement of the reputation system. The results of the experiments partially support that hypothesis: the distributions of buyers become more equitable, but the BLC of the distributions of fair sellers does not vary significantly. We attribute this result to the chosen simulation scenario. Varying the probability of negative reports had an impact on a seller's reputation, but we have not simulated unfair buyers, so no effect on the utilities of fair sellers has been observed.

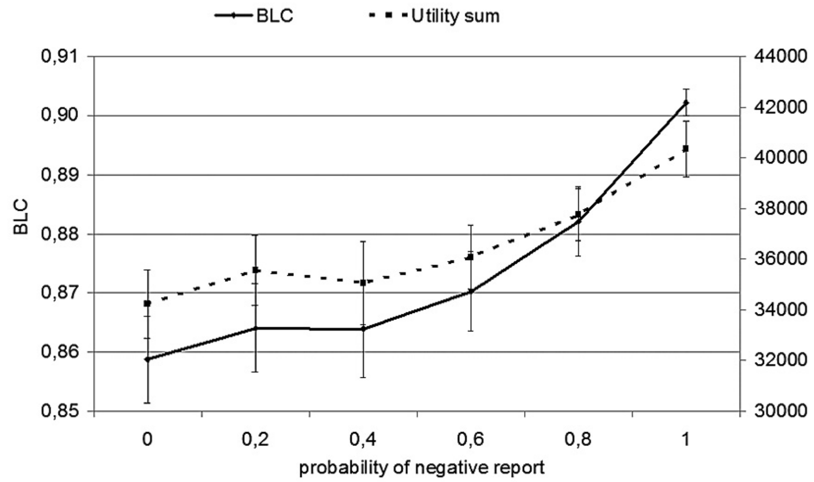


Fig. 3.13. Fairness emergence in the open system

Figure 3.13 shows the effect of increasing the probability of negative reports on the sum of utilities of all buyers and on the BLC of the distribution of buyers' utilities. Since both BLC and the utility sum increase, it can be concluded that the distribution of buyer's reputation indeed becomes more equitable. The effect becomes statistically significant for an increase of the probability of negative reports from 0 to 0.8. Thus, the FE hypothesis is partially confirmed in a realistic open system.

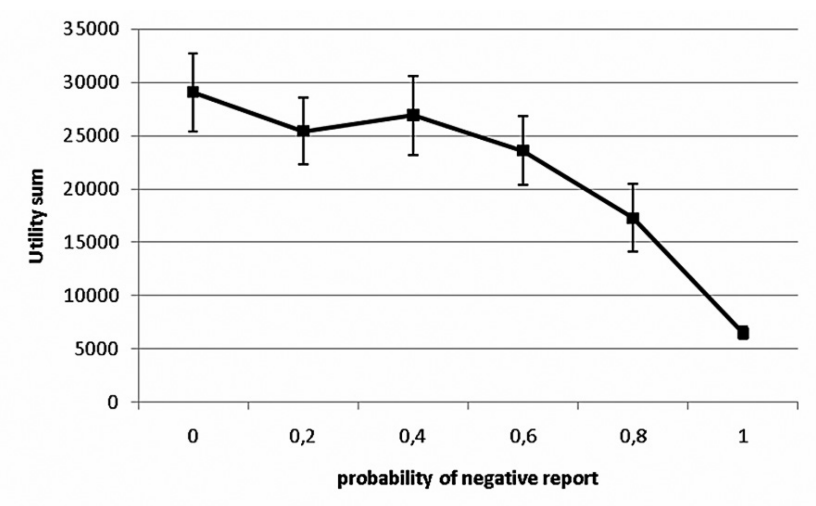


Fig. 3.14. Utilities of unfair sellers in the open system

The reputation system effectively prevents unfair sellers from exploiting buyers. Figure 3.14 shows the sum of utilities of all unfair sellers that decreases with the increasing probability of negative reports. Once again, the effect becomes statistically significant for an increase of the probability of negative reports from 0 to 0.8

We have investigated the sensitivity of Fairness Emergence to the behaviour of unfair sellers (adversaries). To observe this effect, the probability of cheating by a unfair seller was varied. The results are shown on figure 3.15. As expected, the Fairness Emergence was strongest forth case when unfair sellers cheated with probability 1. This meant that the reputation system could easily spot adversaries. Decreasing the probability of cheating weakens Fairness Emergence, with the values of 0.5 as a threshold. For lower probabilities of cheating, Fairness Emergence was not observed. This result can be explained by the simple reputation algorithm used in our simulation scenario (recall that reputation is a simple ratio of the number of fair transactions to the number of all transactions).

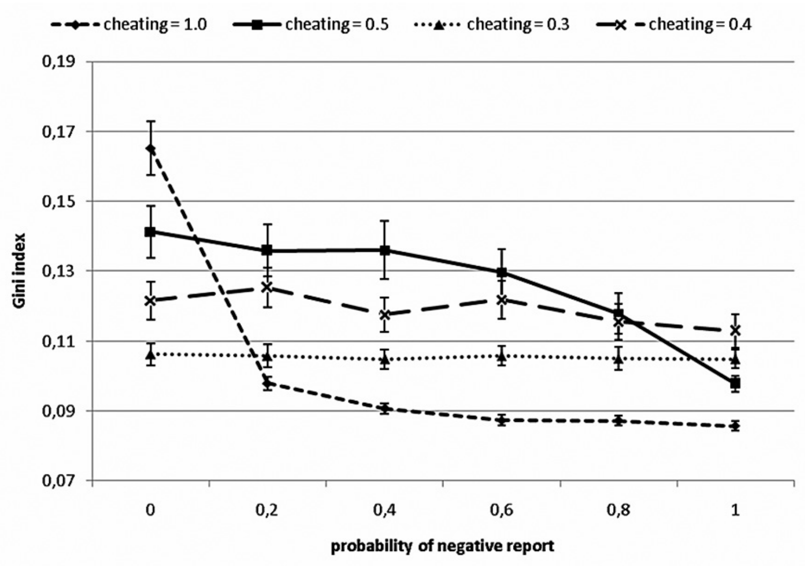


Fig. 3.15. Sensitivity of Fairness Emergence to unfair seller behavior

Table 3.1. Comparison of simple and discriminating adversary strategies for a probability of negative reports of 0.8

	BLC buyers	Utility sum buyers	Utility sum dishonest sellers	Utility sum honest sellers
Simple strategy	0.8461	27,781	30,674	31,182
Discrimination strategy	0.7780	28,361	31,358	30,184

Another, sophisticated adversary behaviour is *discrimination*. An adversary seller that uses a discrimination strategy will cheat only a selected minority of buyers. His selection strategy could be based on the number of transactions that a buyer participated in: the discriminating seller would cheat only inexperienced buyers. In contrast, a simpler adversary would cheat all buyers with an equal probability. Discriminating sellers are harder to detect by the reputation system, because the information about them comes from a minority of buyers. Our simulations have shown that when discriminating strategies are used, fairness emergence is no longer statistically significant. Moreover, the differences in the total sum of utilities are also not statistically significant, while there is a strong increase in the Gini coefficient when a discriminating strategy is used instead of a simple adversary strategy. This result demonstrates the need of explicit consideration for fairness in the evaluation of reputation and TM systems. A comparison of results for discriminating and simple adversary strategies is shown on table 3.1.

More advanced types of reputation algorithms that attempt to recursively weigh received reports with the reputation of reporting agents would be even more vulnerable to discrimination strategies. These types of algorithms, proposed frequently in the literature (Kamvar 2003; Guha 2004), would erroneously overvalue the reputation of discriminating agents, if these act unfairly only towards a minority of discriminated agents.

3.1.7. Conclusion

The Fairness Emergence hypothesis may be viewed as a theoretical concept that is similar to the well-known "evolution of cooperation". On the other hand, it has an inherent practical value. First, if the FE hypothesis holds, then the problem of ensuring fairness in an open, distributed system without centralized control may have found a practical solution: it would suffice to use a good trust management system in order to provide fairness. Second, if the FE hypothesis would not be true in realistic conditions, then a reputation (or trust management) system would allow the existence of a degree of unfairness between similar agents. Such a situation would be highly undesirable from the point of view of users of trust management systems, leading to a disincentive of their usage.

We have shown that the Fairness Emergence hypothesis applies in realistic conditions: in the presence of adversaries and in an unstable state of the system, and also in an open system where the presence of sellers was controlled by a trace from a real Internet auction site. Yet, this work also shows that the FE hypothesis does not apply universally. In particular, fairness emergence does not occur (or is very weak) if very few negative feedbacks are received by the reputation system. The FE hypothesis does not hold if the users of a reputation system are not sufficiently sensitive to reputation or do not have enough choice of transaction partners with a good enough reputation (this implies that if unfair agents would be a large fraction of the population, fairness could not emerge).

Fairness Emergence among buyers was not observed in the open system if the system was not warmed up. The reason for this is that in the open system, some sellers are present only for a few transactions. If the reputation system does not have sufficient information about these sellers, the buyers cannot determine whether they are fair or unfair. It would be possible to initialize the reputations of sellers with a small value, but that would effectively exclude them from the system since it would make it impossible for a new seller to earn a higher reputation. There exists a practical way out of this difficulty: the transactions of new agents could be insured, until their reputation reaches a sufficiently high value. There also exists a practical threat that can lead to a lack of sufficient information about agents: if agents who have a low reputation can assume a new identity (an approach known as whitewashing), then fairness emergence would not occur. This behavior can only be prevented by using stronger authentication of agents.

We have studied the sensitivity of fairness emergence to discrimination attacks. While fairness emergence can still be observed when sellers discriminate a minority of buyers, it is not statistically significant. In simulations when the discriminating agents formed a majority of the population, the FE hypothesis does not hold.

From these results we can draw the following conclusions:

– trust management (reputation) systems can improve distributional fairness in ODS,
– trust management systems should explicitly consider fairness in their evaluation (also in the evaluation of their correctness).

Further research is necessary to establish the sensitivity of the FE hypothesis to more sophisticated attacks on reputation systems. Furthermore, it would be desirable to investigate the emergence of fairness in more general trust management systems, for example n systems that make use of risk in decision support. Another possibility would be the use of transaction insurance together with a reputation system. Last but not least, the use of reputation in practical fair procedures would require a redesign of these procedures – in the light of our results, this is a promising direction of future research.

3.2. Emotion Aware Mobile Application[72]

3.2.1. Introduction

Multi-user operating systems for mobile phones which allow easy user switching doesn't exist. It is probably the most striking proof of the fact that mobile phones

[72] This research was published as: R. Nielek and A. Wierzbicki (2010). Emotion Aware Mobile Application. Computational Collective Intelligence. Technologies and Applications. J.-S. Pan, S.-M. Chen and N. Nguyen, Springer Berlin/Heidelberg. 6422: 122–131.

are exclusively personal devices. In opposite to personal computers we don't share one device but, still, we don't go with personalization beyond theme (skin) selection, accounts' settings and some additional features like five most frequently called numbers.

Supporting and extending personalization is one of main goals of context- aware systems and application but most researches have been focused on a physical or social context (Korpipaa 2004). The true personalization can be achieved only by a device which "feels what user feels". Emotion-aware mobile device followed by emotion-aware services and application are a natural next step in context aware researches. It is worth to notice that already in 1998 Dey (Dey 1998) listed emotional state of a user as an example of context information but since then not much progress towards sensing mobile device has been done. It is so mostly because the internal state of a person who use a mobile device was usually omitted as difficult to measure (compare to localization or temperature) and commercially not very useful. The title of the workshop held during MobileHCI '09 conference "Measuring Mobile Emotions: Measuring the Impossible?" (Geven, Tscheligi and Noldus 2009) perfectly describes concerns shared by the majority of the research community. Lack of good objective measures dedicated to emotional states (it is questionable if they can exist at all) and subjectivity component make results difficult to compare and to justify particular, chosen approach or algorithm.

Problems with an evaluation of the proposed solution are not only limited to measuring emotion. Mizzaro et. al. (Mizzaro 2008) claim that even for a localization based services we don't have good benchmarks which allows us to compare the quality of action-event mapping. Interesting approach was proposed by Oulasvirta (Oulasvirta 2004). Instead of looking for objective benchmark to compare different solution author pinpointed the dominance of the human needs and expectation over technology-driven innovation. In the present-day world innovation is driven neither by technology existing in laboratories nor by pure human needs. In applied computer science innovation are motivated by business. Thus a persuasive use-cases and a proof that proposed solution can be developed with use of existing technology are needed to cause a large-scale deployment.

The approach used in this text is closer to a business people viewpoint. Developed application work on ordinary smartphone and are dedicated for Android – a publicly available mobile operating system. EmotionML, markup language proposed by W3C group, is used to tagging emotion and communication between processing and storing layer in framework. Developed solution is validated with use of SMS corpus collected at the Department of Computer Science at the National University of Singapore by volunteers.

Measuring, understanding and, even, influencing emotional state of end-users (especial in mobile environment) can be a basis for a variety of m-commerce services and products ranged from m-healthcare to matchmaking. Thus, in this chapter a framework for emotional aware mobile applications is sketched. As a proof-of-concept two application are created. SMSemoAlerter that analyze the content of incoming SMS and based on them modify information about user's mood. The second application changes the wallpaper of the mobile phone following by the extracted emotional information. In section 3.2.2., collecting, processing and storing of emotional information is discussed. Section 3.2.3. is devoted to the presentation of the two proof-of-concepts. In the last section conclusion are drafted and some of the ideas for future are presented.

3.2.2. Collecting, processing and storing emotional state

3.2.2.1. Introduction

The framework sketched in this section is composed of three separate layers. Each layer is focused around some important functions and, thus, can work at least semi-independent. Moreover, some algorithms and solutions are platform-independent and can be used universally on almost all present mobile devices (the only limitation is computational power). Another functions have to be developed for only one operating system or even a particular mobile phone.

Modern smartphones can (at least theoretically) hear what you hear, see what you see and even read what you read. So, the first step in developing a system feeling their users' emotion is collecting and preprocessing data (e.g. SMSes, background noises etc.) containing information about emotion. Such functionality is pursued by top layer in proposed framework and is strongly device-dependent.

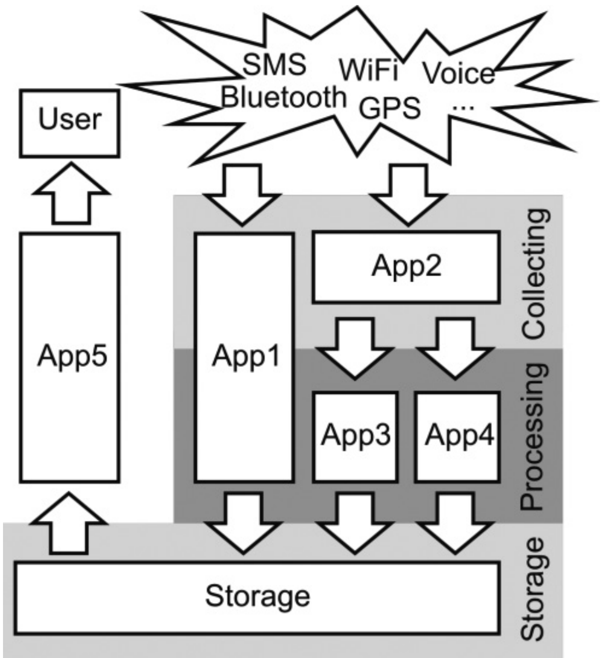


Fig. 3.16. Sketch of the framework with three separate layers

Most promising piece of information are identified, filtered out and forwarded to the next layer for further in-depth processing. Algorithms used on this layer are universal and can be deployed on almost every device regardless of operating system. The bottom layer is responsible for calculating present user's mood based on information incoming from higher layer. Additional functionality of the last layer is affording of collected and computed information for different application.

The proposed framework is presented on figure 3.16. Application developed within this framework are not limited to only one layer and can go across two (quite often) or even three (rare) layers. Services which only use information about user's mood but don't modify them can easily connect to the bottom layer (like App5 on figure 3.16).

3.2.2.2. Collecting facts

A variety of data sources which can deliver some information about user's mood make collecting and preprocessing such data extremely difficult. Necessity to prepare special piece of software not only for every data source but often for many types of mobile devices make the task even more complex. In table 3.2 a selection of possible sources of emotion-rich information in mobile environment is presented. For each data source most promising tools and techniques are given (accompanied by an estimation of required computational power).

Even if for some approaches an efficient and computable algorithms exist it is not always an easy task to gather and processing right data in a real-time. To trace the background noises on a regular basis a microphone embedded into mobile phone has to be active all the time. It will open a lot of question about privacy and decrease the time the device will work on battery to an unacceptable small value. The same problem can be seen more vivid for video stream. On the other hand some emotional rich data like SMSes, email or a history of web browsing are accessible without much e ort. Most mobile operating systems allow developers to register their own classes as a handler of particular event. Then an application is activated by the OS right after such event appeared.

Note that many application can be simultaneous registered to the same event, thus a way to integrate the results of emotion extraction algorithms is required to avoid a situation in which the same emotion could be counted twice. Therefore, the framework presented in this dissertation assume that on the top level only one module is responsible for collecting the data of particular type (let it be video, voice, emails or SMSes) and then, those collected facts are accessible for processing (extracting emotional attitude) on the lower layer.

3.2.2.3. Processing information

Extracting or reasoning emotion from intercepted texts, voice, pictures or other behavioral data is an hard problem which require not only a lot of computational power but also many additional information about context. Constantly growing computational power embedded into mobile devices and intelligent algorithms which help reconstructing proper context are useful but, even then, many unsolvable problems still exist (e.g. subjectivity).

3.2.3. Proof-of-concept

3.2.3.1. SMS EmoAlerter

Dataset

As a basis for validation the proposed solution SMS corpus collected at the Department of Computer Science at the National University of Singapore by volunteers. The corpus, which is distributed under Open Project Library License, contains 10,094 short messages written in English by 133 authors. Average message is composed of 58 characters (standard deviation for the whole dataset is rather big and equal to 39.5 characters). More in-depth analyses and statistics about used dataset can be found in (Korpipaa 2004).

Table 3.2. Sources of emotion rich information in mobile environment

Sources of information	Tools and techniques	Required computational power
SMS, MMS, emails, news (content), calendar	Natural Language Processing tools, dictionaries, Hidden Markov Model and artificial intelligence for disambiguation and misspelling correction	Low to medium for well structured languages; high for languages with complex grammar
Voice tone, background noises	Frequency analyzes, artificial intelligence	High (probably dedicated DSP required)
History of visited web pages	Classification and clustering, predefined categories, artificial intelligence	Low to very low if based on predefined lists of portals
Proximity	Bluetooth, AI for automatic building of databases	Low (energy hungry because of Bluetooth)
Localization	GPS and GSM cells, access points (WiFi)	Low but extensive databases are needed
Video, pictures	Embedded camera, AI (face detection)	Very high
Explicitly declaration	Survey	Very low
Implicitly declaration	Playlist selection etc.	Low

Text processing

The idea to develop a lightweight version of sentiment extraction tool running on mobile phones was driven by the will to verify concepts sketched in previous chapters. Our aim was to prove that mobile devices existing on the market have enough computational power to support NLP-based emotion extraction tool. The usage of special recompiled standalone version of the General Inquirer (Philip 1966) was considered but because of size and computational requirements was rejected. Thus, a simple yet robust model was developed based on the Harvard IV-4 dictionary which contains 1915 positive and 2291 negative words.

In the first step every incoming message is intercepted and tokenized. Because English has very limited flexion stemming is actually unnecessary (except genetivus and plural forms). Thus, a tokenized form can be directly compared with dictionary containing positive and negative words. Typically, as for all samples of user-generated content, also in this dataset a lot of misspellings, mistakes, slangs and abbreviation exist. Because of limited number of characters in which can be sent in one message and lacks of full keyboard in most of used devices many specific abbreviations have evolved and are in common use.

All extracted tokens were compared with a list of English words contains 58 thousands entities. For over 120 thousands tokens found in dataset slightly more than 81 thousands were recognized as an English word. Thus, a need for correcting and decoding mechanism is obvious. Closer look on the 42,646 unrecognized tokens shows that only 3,663 are unique. The most common used unrecognized combination of characters appear in dataset more than 4,000 times and some of them (e.g. "tnk" -> "thank") are rather easy to decode. Therefore for most commonly used abbreviation 31 rules were manually crafted. Selection of prepared rules is presented below: "u" -> "you", "ur" -> "your", "4" -> "for", "n" -> "and", "r" -> "are", "coz" -> "because".

Applying rules described in previous paragraph allows to reduce number of unrecognized tokens to 25 thousands which is less than 20 percent of all tokens in the dataset. It's still higher than in typical texts collected from Internet but as can be expected SMSs contain more words such as the names of persons, organizations, locations or expressions of times because of their informative function.

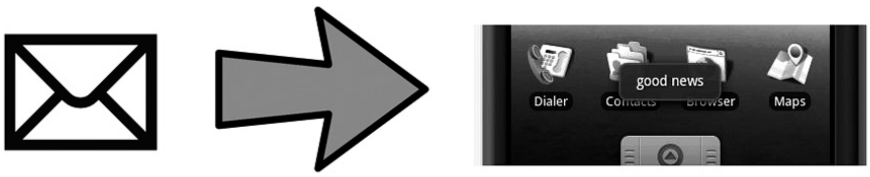


Fig. 3.17. Screenshoot of the SMS EmoAlerter application

Implementation to verify an applicability of the proposed framework a proof-of-concept was developed. As a platform for deployment a platform composed of

T-mobile G1 smart phone and Android v.1.2 operating system was chosen. Currently, it is at best a medium range platform and much more sophisticated devices and operation systems are already on the market (e.g. iPhone 3GS, HTC HD2, Motorola Droid or Google Nexus). Thus, a similar deployment would be also possible, with very limited problems, on cutting the edge mobile phones.

The SMS EmoAlerter remains the architecture of the App1 presented on figure 3.16. Module responsible for collecting facts (in this case it means intercepting incoming short messages) and NLP processing engine (emotion discovering from text) are placed within one application. To intercept incoming SMSes a standard solution delivered by operating system was used. An application was registered as an event handler for the standard event exiting in Android OS which is called RECEIVE SMS. The security policy of the operating system requires that users will explicitly grant this privilege to the application during installation process. When the scheduled event appears, operating system activates application and passes control to it.

Tests

A subset of the dataset described more detailed in the subsection Dataset was manually tagged by two persons. Every single SMS was evaluated either as positive or neutral or negative. The meaning of the SMS was discovered only by looking on their text content. Therefore all additional information like details about sender or history of communication were hidden. The motivation to make the task more complicated for people were twofold. Firstly, we wanted to eliminate as many collateral effects[73], which could influence the human evaluation, as possible. Secondly, to make results of automatic classification and human tagging comparable we had to ensure that this classification will be based on the same dimensions.

The inter-agreement between two persons was relative high and exceed 90% what indicate that negative or positive emotion appeared in SMSes, at least at some level, can be treated as objective feature. The remaining 10% of SMSes in which evaluation of two persons do not match was removed from further process- ing. In table 3.3 the percentage of positive, negative and neutral SMS, according to human evaluation, is presented. The most striking conclusion from that results is that huge majority of messages belong to the neutral category. In overwhelming number of cases a content of SMS was very hard (or even impossible) to understand without precedent messages. Sender of an message, motivated by the limitation of SMS technology and clunky keyboard, usually tries to predict a level of knowledge of receiver and, thus, restricts an content of their message to a new facts which can be placed on the top of the knowledge they share.

[73] It is a well-known phenomena that human usually cannot ignore side-information during evaluation process. Such information like nick or even a combination of num- bers presented next to the content of SMS could influence the emotional state of the evaluator.

Table 3.3. Manually vs. automatic tagged SMSes

	Positive	Neutral	Negative
Manually tagged	15%	81%	4%
Automatic tagged	12%	84%	4%

The limited precision of the NLP algorithm (in comparison to people) in detecting positivity in messages is can be attributed to the limitation of the dictionaries. Standard Harvard IV-4 dictionary contain neither emoticons like ":)" or ";)" nor short combination of letters mimicking the sound of laugh. These two drawbacks can be easily eliminated but the advantage will be very limited as long as we do not start to deal with the history of communication rather than with a single message.

3.2.3.2. Sensing wallpaper

Second example of an application, which use an information about emotional state of user, is the Sensing Wallpaper. Going back to the schema presented on figure 3.16 this application has a design similar to the "App5". It uses information from the storage layer to adapt current wallpaper to the mood of the user. In general two strategies can be identified either the application can improve user's mood by positive wallpapers every time when some bad events (information) affect or the application can follow user's mood and change wallpaper to positive only when the user is happy. The event-action mapping in a context-aware mobile systems is a general and known problem and it is not the aim of this chapter to address it. Therefore, the Sensing Wallpaper application was designed to be agnostic in terms of event-action mapping. Every user can freely change picture-emotional state matching by themselves.

3.2.4. Conclusion

In previous section the framework for developing emotional aware mobile application was presented. Two developed application proof that, firstly, this framework is practically usable and, secondly, that mobile device which is aware of emotional state of their user can be developed with use of software and hardware already present on the market. Sufficient precision of emotion extraction from short messages (SMSes) can be ensured with relative simple algorithm which, at least for English, can run smoothly on middle range phone with Android OS (G1). On the other hand more complex algorithms and another sources of emotion-rich information are still too computational power hungry to be successfully deployed on mobile devices.

Quite surprising was a very low level of emotionality of SMSes in using dataset. The most probably explanation of this fact is the way the dataset was constructed. Volunteers contributing to this dataset could either change their typical communication habits or filter messages that were too emotional. Disregard of the reason, in our research SMSes have appeared to be a tool for exchanging information and agreeing common activity rather than channel for expressing emotion.

Popularization of precise and efficient algorithms for sensing user's emotion and installing them by default on mobile devices can raise some serious privacy and security concerns. The threat that some application will abuse information about our emotional state cannot be fully dismissed. A vivid example of such abuse can be a situation in which service provider (let it be bank or mobile network operator) shapes their offer after our mood or presents it only when we are happy (it's well known psychological effect that we are more willingly to accept proposition/spend money when we are happy). This problem can be partly solved by a mechanism which is present in almost all operating systems for mobile devices and requires explicit declaration of privileges for each application but, firstly, most users are unaware of threats and don't understand the consequences of choices (in matter of security) and, secondly, for devices which are permanently on-line and can be freely managed, inspected and even switched of remotely the true security and privacy actually do not exist. Future research, among security and privacy should be focused on new algorithms, which allow extracting emotion from variety of sources (e.g. video, voice or proximity), and will efficient work with limited computational resources accessible on mobile devices. Yet another interesting topic is looking for practical use-cases for emotion aware application (e.g. monitoring stress factors of user), which will probably raise many additional research questions.

4. Conclusions

This dissertation begins with a question as to whether algorithms can influence people's behavior and if so, how to design algorithms which would support the realisation of social goals. Therefore, the main focus of the preceding chapters is concerned with an examination of social phenomena in behavioural data and developing algorithms which influence the way people take decisions, or their mood. This may leave the question of whether the obtained results justify the affirmative answer to the thesis. There is not, and probably never will be, a tool to categorize all social phenomena because new ways of interacting and constantly changing context (both technological and social) will create a place for new social phenomena to appear. Therefore, giving a universal response to questions posed in the introduction of this dissertation is difficult but making the expectation more realistic and focusing on more practical aspects may allow us to draw some useful conclusions.

In-depth studies of the reputation system in Allegro show that the complex relationships between different users' behavioral patterns and barely predictable side-effects of a particular system design can discourage website owners from a purposeful modification of e-commerce systems. On the other hand, the existence of the spiral of hatred effect is visible proof that the design of algorithms can affect the quality of services available to users, and simple financial instruments (e.g. insurance, escrows or discounts) do not always make up for the bad design of algorithms. The approach proposed in chapter 2.1. and based on developing an external tool for supporting the realisation of social goals is a viable option but even more promising an approach is to convince managers to take a risk and add social goals to their requirements during designing and developing new socially-centric algorithms and systems.

Making the risk associated with changing a complex service used by many people manageable requires delivering answers to a number of questions and concerns. The outcome of the proposed changes should not be either guessed or based on intuition. On the contrary, results should be predicted and all possible changes in people's behaviour at both the micro and macro level (special attention should be paid to emergent phenomena) have to be identified. Social simulations are a very promising tool for modelling and thus, predicting the evolution of complex social systems. In chapter 3.1. simulations of the auction house's reputation system were conducted and their usefulness was proven. Moreover, using data from real system and flexible modelling of adversaries' strategies make the developed simulator realistic and well rooted in reality. Additionally, experiments show that many algorithms can be tested on a small-scale in comparison with a real system; so, multi-million-dollar investments in computational power are not needed.

What was viable as scientific research is not always practical in business. Lack of standardized tools for social simulation next to the deficit of knowledge are the most important barriers in testing all, even small changes with the help of social simulation. Additionally, the need to develop a new model virtually from scratch on each occasion boosts costs and risks. Therefore, considerable work needs to be devoted to:

– preparation of ready-made components of typical social activities in Internet services which can be used "off the shelf",
– developing a library of predefined adversaries' strategies that take into account economic considerations and use the opportunity of cooperation.

Education of managers and customers should not be neglected but as long as convenient tools are not easily available, the rapid development of a practical application of social simulation is not probable.

Reducing risk is important, but it is only one consideration. Another one is to increase the potential profits from implementing algorithms, which takes into consideration social goals. Neither definition of social goals nor the exact identification of profits is an easy task. Social goals are usually expressed in natural language and refer to abstract terms like security, fairness, and justice. Therefore, the very first step which has to be done is a transformation of fuzzy, linguistic concepts into well-defined, computable formulas. In this dissertation the theory of equity was used as a starting point for constructing a formal, measurable definition of fairness (a detailed description can be found in section 3.1.2). However, this particular solution cannot be applied to all social goals, the proposed approach is all-purpose and can be adapted to express a broad variety of goals.

Security or fairness is a universally desirable property and, excluding adversaries, most users of information systems want to feel more secure and not discriminated. At the same time, owners of e-commerce systems (e.g. auction houses or shops) are focused on the increasing turnover and, which is usually mutually connected, profits. These goals do not need to be contradictory to each other, so a well-designed algorithm should simultaneously take them into account. As a proof-of-concept that such mix of objectives is doable, a multi-criteria function which simultaneously evaluates an equitable distribution of goods and the aggregate number of transactions has been designed. The proposed new reputation algorithms increase the security of transactions but at the same time do not handicap the total number of transactions, thus are very easy to accept for e-commerce sites' owners.

Looking for social phenomena in reputation systems and creating algorithms for supporting equitable distribution in auction houses proves that at least for centralized, homogenous systems theses of this dissertation are true. The results obtained for attempts of prediction of shares prices on the Warsaw Stock Exchange show that heterogeneity is not an insurmountable obstacle. Sensing social phenomena is possible also in weak (to non) structured data collected from the Internet. Lack of authorship, only rough estimation of the number of visitors and noisy datasets make the task more complicated but still doable. Moreover, in heterogeneous systems not only sensing but also influencing is possible. The framework developed in chapter

3.2. delivers examples how mobile phones can be used to collect information about people's mood and how people's mood can be influenced by the information system. Algorithms trying to increase wellbeing in society can use such a framework.

The idea that algorithms can be developed to assure the selected social goals even in a variable social context can find a very broad implementation. Electronic intermediaries in communication are a very promising area of future research. Clever algorithms can help to maintain valuable relationships by suggesting partners and subjects to mention. The Tlenoklika, a plugin to the one of the most commonly used Polish instant messengers, focused on discovering the nature of the relations between people and refreshing fading relationships, shows that it is not a purely theoretical possibility (Doniec, Hupa and Nielek 2009). A similar approach can also be used to limit an unrestricted growth of number of "facebook" friends. Defriending can be seen as a new social phenomenon, which is also followed by an extension of English dictionariess[74]. Algorithms supporting the realisation of social goals can be useful not only in the rapidly developing new systems but also in the already existing and mature ones. Carefully selecting and recommending appropriate team members can significantly improve the quality of articles on Wikipedia. Research published by (Wierzbicki, Turek and Nielek 2010) is very promising and confirms that this can be realized in practice, based only on the edition of history and talk pages.

More and more technologies are developed to be inextricably connected with social context. Such technologies are a potential source of many different social phenomena which should be a priori identified and modified to maximize the benefits of users. This dissertation has proven that the design of algorithms and information technology can take into account social goals. In subsequent years we will probably see the flowering of social-centric technologies and algorithms which will eventually not only improve our experiences with information technology but also deliver a huge amount of data for researchers.

[74] Process of elimination of links to people on social networking sites (mainly on the Facebook) is called either unfriending or defriending. The first word has been selected as a 2009 Word of the Year by Oxford American Dictionary, the second one seems to be more frequently used by Facebook users.

Bibliography

Aone C., Okurowski M.E., Gorlinsky J., *Trainable, scalable summarization using robust NLP and machine learning*, Proceeding COLING ,98 Proceedings of the 17th international conference on Computational linguistics, Association for Computational Linguistics, Montreal 1998, 62–66.

AP. *Max's privacy war brings Facebook to heel 27 October 2011*, The Age; http://www.theage.com.au/technology/technology-news/maxs-privacy-war-brings-facebook-to-heel-20111027–1mksg.html (2.03.2012).

Axelrod R., *The Evolution of Cooperation*, Basic Books, New York 2006.

— *Advancing the art of simulation in the social sciences*, Complexity, 1998.

Baayen H., *Word frequency distributions*, Springer 2001.

Bajari P., Hortaçsu A., *The Winner's Curse, Reserve Prices, and Endogenous Entry: Empirical Insights from eBay Auctions*, The RAND Journal of Economics, Vol. 34, No. 2 (Summer 2003), 329–355.

Bar-Yossef Z., Gurevich M., *Estimating the impresionrank of web pages*, Proceedings of the 18th Interantion Conference on World Wide Web, New York: ACM 2009, 41–50.

Bedau A.M., Humphreys P., *Emergence. Contemporary Readings in Philosophy and Science,* MIT Press 2007.

Beltran F.S. et al., *Forecasting a Language Shift Based on Cellular Automata*, Journal of Artificial Societies and Social Simulation n. 12 (3) 5, 2009.

Berners-Lee T., Hall W., Hendler J.A., O'Hara K., Shadbolt N., Weitzner D.J., *A Framework for Web Science*, Foundations and Trends in Web Science (now Publishers Inc.) 1, No. 1 (2006), 1–134.

Berners-Lee T., Hall W., Hendler J., Shadbolt N., Weitzner J., *Computer Science: Enhanced: Creating a Science of the Web*, "Science" 2006, 769–771.

Berriman B., Groom S.L., *How Will Astronomy Archives Survive the Data Tsunami?*, Queue 2011, No. 9, 21–29.

Bolc L., Makowski M., Wierzbicki A., *Preface,* Social Informatics, Springer 2010.

Botsch J., Luckner S., *Empirische Analyse von Bewertungskommentaren des Reputationssystems von eBay*, Multikonferenz Wirtschaftsinformatik 2008, 1199–1210.

Breznitz D., Murphree M., Goodman S., *Discussion about privacy*, "Computer Innovative Technology for Computer Professionals 44" 2011, No. 6, 100–102.

Brodkin J., *French court frowns on Google autocomplete, issues $65,000 fine*, ars technica; http://arstechnica.com/tech-policy/news/2012/01/french-court-frowns-on-google-autocomplete-issues-65000-fine.ars (3.03.2012).

Buczynski A., Przepiorkowski A., *Demo: An Open Source Tool for Partial Parsing and Morphosyntactic Disambiguation*, Proceedings of LREC 2008.

Buczynski A., Wawer A., *Shallow parsing in sentiment analysis of product reviews*, Proceedings of the Partial Parsing workshop at LREC 2008, 14–18.

Cartwright N., *How the Laws of Physics Lie*, Clarendon Press, Oxford 1999.

Castelfranchi C., Conte R., Paolucci M., *Normative Reputation and the Costs of Compliance*, "Journal of Artificial Societies and Social Simulations" 1988, Vol. 1, No. 3.

Chakrabarti S., van der Berg M., Dom B., *Focused crawling: a new approach to topic-specific Web resource discovery*, Proceedings of 8th International World Wide Web Conference (WWW8) 1999.

Cho V., Wüthrich B., Zhang J., *Text Processing for Classification*, "Journal of Computational Intelligence in Finance" 1999, No. 7, 6–22.

Claes R., Holvoet T., *Multi-model Traffic Microsimulation*, Proceedings of the 2009 Winter Simulation conference 2009, 1113–1123.

Crystal D., *Internet Linguistics: A Student Guide*,Routledge, New York 2011.

Czarnecki L., *Designing system for tracking users' behaviours during surfing*, BSc thesis, Polish-Japanese Institute of Information Technology, Warszawa 2012.

Davidsson P., *Agent Based Social Simulation: A Computer Science View*, "Journal of Artificial Societies and Social Simulation" 2001, Vol. 5, No. 1.

Decoster A., Loughrey J., O'Donoghue C., Verwerft D., *Microsimulation of indirect taxes*, "International Journal of Microsimulation" 2011, 41–56.

Dellarocas C., *Immunizing Online Reputation Reporting Systems Against Unfair Ratings and Discriminatory Behavior*, Proc. of the 2nd ACM Conference on Electronic Commerce, Minneapolis 2000, 150–157.

Demetris A. et al., *we.b: the web of short urls*, Proceedings of the 20th international conference on World wide web, ACM, New York 2011, 715–724.

Deutsch M., *Experimental studies of the effects of different systems of distributive justice*, in: *Social comparison, social justice, and relative deprivation*, Lawrence Erlbaum, Hillsdale 1987, 151–164.

— *Equity, equality, and need: What determines which value will be used as the basis of distributive justice?*, "Journal of Social Issues" 1975, 137–149.

Dey A.K., *Context-aware computing: The CyberDuck project*, Spring Symposium on Intelligent Environments, AAAI Press 1998, 51–54.

Dietz J., *Extreme breakthrough*, "Teradata Magazine Online", 1 September 2009; http://www.teradatamagazine.com/v09n03/Viewpoints/Extreme-breakthrough/(1.03.2012).

Dietz K., Heesterbeek J.A.P., *Daniel Bernoulli's epidemiological model revisited*, "Mathematical Biosciences" 2002, 1–21.

Doctorow C., *Tracking the astounding pace of digital storage*, boingboing 2011; йил 8-March; http://boingboing.net/2011/03/08/tracking-the-astound.html (1.03.2012).

Doniec A., Hupa A., Nielek R., *Web of Friends – Discovering a Social Network by Mining Data from Instant Messengers*, 2009 International Workshop on Social Informatics, IEEE 2009, 16–22.

Ed., *Sensors and sensitivity*, "The Economist", 4 June 2009; http://www.economist.com/node/13725679 (2.03.2012).

Edmonds B., *Towards an Ideal Socia lSimulation Language*, Springer-Verlag 2003.

Elgesem D., *Normative Structures in Trust Management*, Trust Management, Springer 2006.

Epistemological Aspects of Computer Simulation in the Social Sciences, Second International Workshop EPOS, Bresca: Revised Selected and Invited Papers 2006.

Epstein J.M. *Generative Social Science: Studies in Agent-Based Computational Modeling*, "Princeton University Press" 2007.

— *Agent-based computational models and generative social science*, Complexity 1999, 41–57.

Evaangelopoulos, N., Visinescu L. "Text-Mining the Voice of the People." *Communications of the ACM* 55 (February 2012), 62–69.

Feurbaey M., *Fairness, Responsibility and Welfare,* "Oxford University Press" 2008.

Friedman E.J., Resnick P., *The Social Cost of Cheap Pseudonyms*, "Journal of Economics and Management Strategy" 2001, No. 10, 173–199.

Ganesan K.A., Sundaresan N., Deo H., *Mining tag clouds and emoticons behind community feedback*, In Proceeding of the 17th international Conference on World Wide Web. Beijing, China: ACM 2008, 1181–1182.

Gantner Z., Schmidt-Thieme L., *Automatic content-based categorization of Wikipedia articles*, Proceedings of the 2009 Workshop on The People's Web Meets NLP, Stroudsburg: ACM 2009, 32–37.

Geven A., Tscheligi M., Noldus L., *Measuring Mobile Emotions: Measuring the Impossible?*, Proc. of the 11th Int. Conference on Human-Computer Interaction with Mobile Devices and Services 2009, 109:1–109:3.

Gilbert N., *Artificial societies*, University of Surrey 1991.

Gilbert N., Troitzsch K.G., *Simulation for the Social Scientist*, Open University Press, New York 1999.

Gordijn J., Akkermans H., *Designing and evaluating e-Business models*, IEEE Intelligent Systems 2001, 11–17.

Gouws S. et al., *Contextual bearing on linguistic variation in social media*, Proceedings of the Workshop on Languages in Social Media, New York: Association for Computational Linguistics 2011, 20–29.

Granger C.W.J., *Investigating causal relations by econometric models and cross-spectral methods*, "Econometrica" 1996, No. 37 (3), 424–438.

Gregg D.G.,Scott J., *A Typology of Complaints About Ebay Sellers*, Communications Of The ACM 51:4 (April 2008), 69–74.

Grimm V., Railback S.F., *Individual-based Modeling and Ecology*, Princeton University Press 2005.

Grosseck G., *Repere Indentitareprivind Informatica Sociala, Revisita de Informatica Sociala*, Report to the Norwegian Parliament 1985, 32–38.

Guha R., Kumar R., Raghavan P., Tomkins A., *Propagation of trust and distrust*, Proceedings of the 13th international conference on World Wide Web 2004, 403–412.

Gurevich M., Ziv B.-Y., *Estimating the impressionrank of web pages*, Proceedings of the 18th international conference on World Wide Web, New York: ACM 2009, 41–50.

Hendler J., Shadbolt N., Hall W., Berners-Lee T., Weitzner D., *Web science: an interdisciplinary approach to understanding the web*, Commun. ACM 2008, 60–69.

http://hmi.ucsd.edu/pdf/HMI_2009_ConsumerReport_Dec9_2009.pdf.

http://petewarden.typepad.com/searchbrowser/2010/04/how-i-got-sued-by-facebook.html.

http://www.theage.com.au/technology/technology-news/maxs-privacy-war-brings-facebook-to-heel-20111027–1mksg.html.

Humphreys P., Bedau M.A., *Emergence. Contemporary Readings in Philosophy and Science*, "The MIT Press" 2007.

Instructions per second, Wikipedia; http://en.wikipedia.org/wiki/Instructions_per_second.

Isomursu M., Tahti M., Vainamo S., Kuutti K., *Experimental evaluation of five methods for collecting emotions in field settings with mobile applications*, J. Human-Computer Studies 2007, 404–418.

Izquierdo L.R., Polhill J.G., *Is Your Model Susceptible to Floating-Point Errors?*, "Journal of Artificial Societies and Social Simulation" 2006, Vol. 9, No. 4.

Kamvar S.D., Schlosser M.T., Gracia-Molina H., *The EigenTrust Algorithm for Reputation Management in P2P Networks*, Proceedings of the Twelfth International World Wide Web Conference 2003, 640–651.

Karanam S., van Oostendorp H., Indurkhya B., *The role of content in addition to hyperlinks in user-clicking behavior*, Proceedings of the 28th Annual European Conference on Cognitive Ergonomics, New York ACM 2010, 125–131.

Khademhoseiny M., *The sociology of Internet shaping*, Proceedings of the WebSci10: Extending the Frontiers of Society On-Line, Raleigh 2010.

Khopkar T., Li X., Resnick P., *Self-selection, slipping, salvaging, slacking, and stoning: the impacts of negative feedback at eBay*, ACM Conference on Electronic Commerce, ACM 2005, 223–231.

Kirkpatrick D., *The Facebook Effect*, Simon & Schuster 2010.

Kleinberg J., *The convergence of social and technological networks*, Communications of the ACM 51 (11) 2008, 66–72.

Kling R., *What is Social Informatics and Why Does it Matter*, "*D-Lib Magazine* 5", No. 1 (January 1999).

Korpipaa P., Hakkila J., Kela J., Ronkainen S., Kansala I., *Utilising context ontology in mobile device application personalisation*, Proceedings of the 3rd international conference on Mobile and ubiquitous multimedia, ACM 2004, 133–140.

Kostreva M., Ogryczak W., *Linear optimization with multiple equitable criteria*, RAIRO Operations Research 1999, 275–297.

Kostreva M., Ogryczak W., Wierzbicki A., *Equitable aggregations and multiple criteria analysis*, European Journal of Operational Research No. 158, 2004, 362–377.

Kurosu T., *In Search of a Paradigm of Socio-Informatics: On Socio-Informatics and Social informatics*, "Journal of Socio-Informatics" 2010, 69–81.

Lamb R., Sawyer S., *On extending social informatics from a rich legacy of networks and conceptual resources*, "Information Technology & People" 2005, No. 3, 9–20.

Lavrenko V., Schmill M.O., Lawrie D., Ogilvie P., Jensen D., Allan J., *Mining of Concurrent Text and Time Series*, Proceedings 6th ACM SIGKDD Int. Conference on Knowledge Discovery and Data Mining, Boston 2000, 37–44.

Lawrence S., Giles C.L., *Searching the world wide web*, "Science" 1998, 98–100.

Lee S., Sherwood R., Bhattacharjee B., "Cooperative peer groups in NICE" 2003, No. 2, 1272–1282.

Lee Z., Im I., Lee S.J., *The Effect of Negative Buyer Feedback on Prices in Internet Auction Markets*, The Proceedings of the 21st International Conference on Information Systems, (ed.) W. Orlikowski et al., 2000, 286–287.

Lin Z., Li J., *The online auction market in China: a comparative study between Taobao and eBay*, ICEC 2005, 123–129.

Lissowski G., *Zasady sprawiedliwego podziału dóbr (Principles of Fair Distribution of Goods)*, Wydawnictwo Naukowe Scholar, Warsaw 2008.

Liu L., Zhang S., Dong Ryu K., Dasgupta P., *R-Chain: A Self-Maintained Reputation Management System in P2P Networks*, Proc. ISCA PDCS, 2004, 131–136.

Lucking-Reiley D., Bryan D., Prasa N., Reeves D., *Pennies from eBay: The Determinants of Price in Online Auctions*, "The Journal of Industrial Economics" (Willey) 2007, July, 223–233.

Malaga R.A., *Web-based reputation management systems: Problems and suggested solutions*, "Electronic Commerce Research" 2001, 403–417.

Menezes A., *Better Contextual Translation Using Machine Learning*, Proceedings of the 5th Conference of the Association for Machine Translation in the Americas on Machine Translation: From Research to Real Users, Springer-Verlag 2002, 124–134.

Merritt R., *Nokia tweaks Bluetooth for indoor navigation*, "EE Times" 29 November 2011; http://www.eetimes.com/electronics-news/4230993/Nokia-tweaks-Bluetooth-for-indoor-navigation (3.03.2012).

Milian M., *Why text messages are limited to 160 characters*, "Los Angeles Times" 2009 3-May; http://latimesblogs.latimes.com/technology/2009/05/invented-text-messaging.html (1.03.2012).

Miller J.M., *How Nokia's 41-Megapixel Smartphon Works*, Forward Thinking, 27 February 2012; http://forwardthinking.pcmag.com/show-reports/294650-how-nokia-s-41-megapixel-smartphone-works (2.03.2012).

Mittermayer M., Knolmayer G., *Text Mining Systems for Market Response to News: A Survey*, University of Bern, Institute of Information Systems, Bern 2006.

Mizzaro S., Nazzi E., Vassena L., *Retrieval of context-aware applications on mobile*, Information in Context 2008, 65–71.

Mody R.N., Kerstein K.S., Willis R., *WiMo: location-based emotion tagging*, Proceedings of the 8th International Conference on Mobile and Ubiqui-, ACM 2009, 1–4.

Montanola-Sales C., Onggo B.S., Casanovas-Garcia J., *Agent-based simulation validation: A case study in demographic simulation*, SIMUL 2011: The Third International Conference on Advances in system Simulation 2011, 101–107.

Morzy M., Wierzbicki A., *The Sound of Silence: Mining Implicit Feedbacks to Compute Reputation*, WINE, Springer, Berlin 2006, 365–376.

Morzy M., Wierzbicki A., *The Sound of Silence: Mining Implicit Feedbacks to Compute Reputation*, Proc. 2nd international Workshop on Internet & Network Economics, Springer, Berlin 2006, 365–376.

Mowshowitz A., *The Conquest of Will,* Reading: Addison-Wesley 1976.

Mui L., *Computational Models of Trust and Reputation: Agents, Evolutionary Games, and Social Networks (Ph.D. Dissertation)*, Massachusetts Institute of Technology 2003.

Nakanishi Y., Kitaoka N., Ohyama M., Hakozaki K, *Estimating Communication Context throught Location Information and Schedule Information*, CHI 2002, 642–643.

Nam sued for libel over comments on eBay; http://www.telegraph.co.uk/news/uknews/3247683/Man-sued-for-libel-over-comments-on-eBay.html.

Nuno D., *Validation and Verification in Social Simulation: Patterns and Clarification of Terminology in Epistemological Aspects of Computer Simulation in the Social Sciences*, "Computer science" 2009.

Ogryczak W., *On Principles of Fair Resource Allocation for Importance Weighted Agents*, Proc. First International Conference on Social Informatics, Warsaw 2009, 57–62.

Oreskes N., Shrader-Frechette K., Belit J., *Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences*, "Science" 1994, Vol. 263, 641–648.

Oulasvirta A., *Finding Meaningful Uses for Context-Aware Technologies: The Humanistic*, CHI 2004, 247–254.

Ozakca M., Lim Y., *A study of reviews and ratings on the internet*, CHI ,06 Extended Abstracts on Human Factors in Computing Systems, Montréal–Québec, ACM 2006, 1181–1186.

Philip J., *The General Inquirer: A Computer Approach to Content Analysis*, The MIT Press 1966.

Polhill J.G., Izquierdo R.L., Gotts N.M., *The Ghost in the Model (and Other Effects of Floating Point Arithmetic)*, "Journal of Artificial Societies and Social Simulation" 2005, vol. 8, no. 1.

Pollock G.B., Dugatkin L.A., *Reciprocity and the Evolution of Reputation*, "Journal of Theoretical Biology" 1992, 25–37.

Przepiorkowski A., Kupsc A., Marciniak M., Mykowiecka A., *Formalny opis języka polskiego. Teoria i implementacja*, Akademicka Oficyna Wydawnicza, Warszawa 2002.

Raento M., Eagle N., *Smartphones: An Emerging Tool for Social Scientists*, "Sociological Methods Research" 2009, No. 37:3, 426–454.

Rawls J., *The Theory of Justice*, Harvard Univ. Press, 1971.

Repast; Repast Organization for Architecture and Development 2003.

Resnick M., *Turtles, Termites, and Traffic James. Explorations in Massively Parallel Microworlds*, The MIT Press 1997.

Resnick P., Zeckhauser R., *Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System*, "The Economics of the Internet and E-Commerce" 2002, No. 11, 127–157.

Rossiter S., Noble J., Bell K.R., *Social Simulations: Improving Interdisciplinary Understanding od Scientific Positioning and Validity*, "Journal of Artificial Societies and Social Simulation" 2010, 13 (1) 10.

Roy, D., et. al. *The Human Speechome Project*, Proceedings of the 28th Annual Cognitive Science Conference. 2006.

Rungsawang A., and N. Angkawattanawit, *Learnable topic-specific web crawler*, Journal of Network and Computer Applications, 2005, 97–114.

Ryaboy, D., *Which companies produce more than 10TB data per day?*, Quora; http://www.quora.com/Which-companies-produce-more-than-10TB-of-data-per-day (2.03.2012).

Sanchez J.M., Cano J.C., Calafate C.T., Manzoni P., *BlueMall: A Bluetoothbased Advertisement System for Commercial Areas*, PM2HW2N '08, Vancouver 2008, 17–22.

Sawyer R.K., *Social Emergence. Societies as Complex Systems*, Cambridge University Press, 2005.

Scheff T., Fearon D., *Social and Emotional Components in Self-Esteem*, "Journal of the Theory of Social Behavior" 2004, 73–90.

Schelling T.C., *Micromotives and Macrobehavior*, W.W. Norton 1978.

Searle J., *Reductionism and the Irreducibility of Consciousness*, in: *Emergence. Contemporary Readings in Philosophy and Science*, (eds.) P. Humphreys, M.A. Bedau, MIT Press, 2008.

Sen A, *Collective Choice and Social Welfare*, Holden Day, San Francisco 1970.

Seo Y., Giampapa J.A., Sycara K., *Text Classification for Intelligent Portfolio Management. Technical Report CMU-RI-TR-02–14*, Robotics Institute, Carnegie Mellon University, Pittsburg 2002.

Shorrocks A, *Ranking Income Distributions*, "Economica" 1983, 3–17.

Sigler M.G., *Eric Schmidt: Every 2 Days We Create As Much Information As We Did Up To 2003*, TechCrunch, 4 August 2010; http://techcrunch.com/2010/08/04/schmidt-data/ (3.03.2012).

Sobkowicz P., *Modelling Opinion Formation with Physics Tools: Call for Closer Link with Reality*, "Journal of Artificial Societies and Social Simulation" 2009, 12 (1) 11.

Standifird S.S., *Reputation and E-commerce: eBay Auctions and the Asymmetrical Impact of Positive and Negative Ratings*, "Journal of Management" 2001, 279–295.

Stauffer D., *Sociophysics: the Sznajd model and its application*, "Computer Physics Communications" 2002, 93–98.

Stoifo S., Bellovin S.M., Evans D. "Measuring Security." Security & Privacy (IEEE), 2011, 60–65.

Sznajd-Weron K., *Sznajd Model and its Applications*, Acta Physica Polonica B 2005, Vol. 36, No. 8, 2537–2547.

Talbot D., *A Social-Media Decoder*, "Technology review", 1 November 2011; http://www.technologyreview.com/computing/38910.

Talwar A., Jurca R., Faltings B., *Understanding user behavior in online feedback reporting*, Proceedings of the 8th ACM Conference on Electronic Commerce, New York ACM 2007, 134–142.

Tan Y., Thoen W., Gordijn J., *Modeling controls for dynamic value exchanges in virtual organizations*, Trust Management: Second International Conference (iTrust), Springer, Oxford 2004, 236–250.

Tan Y., Thoen W., Gordijn J., *Social justice and social movements*, in: *The handbook of social psychology*, McGraw-Hill, Boston 1998, 595–629.

Tapscott D., Williams A.D., *Macrowikinomics: Rebooting Business and the World*, Portofolio Hardcover, London 2010.

— *Wikinomics: How Mass Collaboration Changes Everything*, Portfolio Trade, 2006.

Tetlock P. C. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *Journal of Finance*, 2007, 1139–1168.

Troitzsch K.G., *Simulation and Social Sciences: On the Future of a Difficult Relation*, Social Science Microsimulation, 1995, 459–460.

Tsai T., Lo J., Chen W., *Design and Implementation of Mobile Personal Emotion*, Proceedings of the 2009 Tenth International Conference on Mobile Data Management: Systems, Services and Middleware, IEEE Computer Society, 2009, 430–435.

Turkle S., *Simulation and its discontents,* The MIT Press, 2009.

Tyler T.R., Smith H.J., *Social justice and social movements*, in: *The handbook of social psychology*, McGraw-Hill, Boston 1998, 595–629.

van Fraassen B., *The Science Image*, Oxford University Press, 1980.

Vehovar V., *Social Informatics: An Emerging Discipline?*, Social Informatics: An information society for all? In remembrance of Rob Kling, IFIP International Federation for Information Processing, 2006, 73–85.

Wawer A., Nielek R., *Application of Automated Sentiment Extraction from Text to Modeling of Public Opinion Dynamics*, "Polish Journal of Environmental Studies" 2008, Vol. 17, No. 3B, 508–513.

Weizenbaum J., *Computer Power and Human Reason*,W.H. Freeman, San Francisco 1976.

Wesolowski A., Eagle N., *Parametrizing the Dynamics of Slums*, AAAI Spring Symposium 2010 on Artificial Intelligence for Development 2010, 103–108.

Wierzbicki A., *The Case for Fairnessof Trust Management*, Proc. 3rd International Workshop onSecurity and Trust Management (STM 07), Elsevier, 2007, 73–89.

— *Trust Enforcement in Peer-to-peer Massive Multi-player Online Games*, Proc. Grid computing high-performance and Distributed Applications (GADA'06), Springer, 2006, 1163–1180.

Wierzbicki A., Kaszuba T., Nielek R., Datta A., *Trust and Fairness Management in P2P and Grid systems*, in: *Handbook of Research on P2P and Grid Systems for Service-oriented Computing: Models, Methodologies and Applications*, (eds.) G. Exarchakos, M. Li, A. Liotta, N. Antonopoulos, IGI-Global, 2009, 748–773.

Wierzbicki A., Turek P., Nielek R., *Learning about team collaboration from Wikipedia edit history*. Int. Sym. Wikis 2010, Poland, Gdans.

Wilks Y., *Natural Language Processing as a Foundation of the Semantic Web*, Foundations and Trends in Web Science 1, 2010, No. 3–4, 199–327.

Wilson E.O., *Social Evolution*, Benjamin Cummings, 1985.

Wydra H., *The Recurrence of Violence*, "Sociology Compass" 2008, No. 1, 183–194.

Yang B., Lugger M., *Emotion recognition from speech signals using new harmony features*, Signal Processing 2010, Vol. 90, No. 5, 1415–1423.

Young H. P., *Equity: In Theory and Practice*, Princeton University Press 1994.

List of figures

List of tables