WHAT DO CROATIAN SCIENTIST WRITE ABOUT? A SOCIAL AND CONCEPTUAL NETWORK ANALYSIS OF THE CROATIAN SCIENTIFIC BIBLIOGRAPHY

Markus Schatten*

Faculty of Organization and Informatics – University of Zagreb
Varaždin, Croatia

ABSTRACT

This article analyses the Croatian Scientific Bibliography (CROSBI) as a social application using a number of different approaches. By analysing and visualizing the conceptual network the core of keywords is determined for each scientific field: biomedical sciences, biotechnology, social sciences, humanities, natural sciences and technical sciences. Through the interpretation of core concepts by meta-data from another social application (Wikipedia) it is concluded about the disproportion of interpretative capabilities of two social systems: the Croatian scientific community and the public. Additionally through a social network analysis between scientific areas according to a social (scientific collaboration) and conceptual (keyword co-affiliation) another disproportion is revealed regarding interdisciplinary research.

KEY WORDS

CLASSIFICATION

*Corresponding author, η: markus.schatten@foi.hr; +38542390806;
Faculty of Organization and Informatics, Pavlinska 2, HR – 42 000 Varaždin, Croatia

INTRODUCTION

The Croatian Scientific Bibliography[1] – CROSBI (in Croatian: *Hrvatska znanstvena bibliografija*) allowed anyone with Internet connectivity a detailed insight into the production of Croatian scientists. An interesting feature of the CROSBI project, which we want to point out here is that the bibliography is being populated by the authors them self, which makes it a social application in its own right. Hence, it represents a reflection of a social system that can be analysed by analysing the reflection [1, 2]. Particularly, if we define a social system as an autopoietic system [3] in a broader environment (whereby the social application is part of the environment) the process by which the system leaves trails in the environment by transforming it can be viewed as the process of structural coupling [4].

According to this view, the social system exchanges components of its structure with the environment (in our case written communication which is in accordance with the semantics and syntax of the social system). On the other hand, since social systems are meaning processing systems, one might ask the question if it is possible to interpret the left trails by other social systems?

To further justify the approach, consider analyses of two (human) autopoietic systems, Humberto Maturana and Niklas Luhmann. They both structurally coupled with publishing technology (alopoietic systems) and left trails in form of their published writings. That allowed us (and other researchers) to perform an analysis of their thoughts like in [4] by actually interpreting their trails. The analysis allowed us to interpret concepts like structure and organization by using the definitions of one and the other. As it is in detail elaborated in [1], we can use autopoietic theory as a framework to understand complex systems, regardless of their origin (biological, social, information system), an this article is just an invocation of this principle.

In the following research we will try to analyse a social system (the Croatian scientific community) through an attempt of interpreting its trails through another social system (the public). As a reflection of the public as a social system we will use the Wikipedia system, the free on-line encyclopedia, which besides its (most profound) English version has versions for most world languages, including Croatian.

The core of this research is thus a comparative analysis of two social systems through their reflections: the Croatian scientific community (with CROSBI as the reflection) and the public (with Wikipedia as the reflection). The Croatian scientific community will be analysed particularly within for its social and conceptual connections as well. In accordance with that we establish the following two hypotheses: (H1) there is a disproportion between the topics the Croatian scientific community deals with and the understanding of these topics by the public, and (H2) there is a disproportion among the conceptual interconnection between scientific fields and the level of collaboration between them.

We will consider the first hypothesis as confirmed if at least 50 % of core concepts from the Croatian scientific community social system are not interpretable in the reflection of the public social system. The second hypothesis shall be considered confirmed if there is a disproportion between social and conceptual interconnection of at least 50 %. Both of these values are not arbitrary since they indicate misunderstanding in the former, and irrational behaviour in the letter case.

To analyse the semantics (meaning and conceptual interconnection) of a system, a number of numerical and logical methods exist coming from different fields like data-mining and knowledge discovery [5], the semantic Web [6], social Web mining [7, 8], etc. Herein we will focus on the ACI (actor-concept-instance) method which was initially used by Mika to

analyse social applications [9]. In accordance with the needs of the research at hand we will extend the method from tripartite to n-partite graphs, as shall be described in the following.

In the end of this introduction, we need to point out that analysing Croatian scientific bibliography is not a new thing and there are recent studies which provide different advances in the bibliometric and/or scientometric analysis of individual journals [10, 11], scientific subjects [12], scientific fields [13], or even further [14]. Without diminishing the importance of these studies, in this article I will for the first time provide a conceptual and social network analysis of the Croatian scientific bibliography, as well as a visualization of the acquired networks in order to gain insight into the main subjects (concepts) Croatian scientists wrote about in the past decade (2000-2010), as well as understand their mutual interconnections.

METHODOLOGY

Network theory, or, as Barabasi calls it [15], the new science of networks, studies social, biological, transport, technological, physical, semantic and other types of networks. The field of social network analysis has a long tradition, but only after the emergence of the Internet and contemporary information and communication technology it gained a huge impetus.

A network can be defined as a mathematical abstraction which consists of two parts: (1) nodes (which can represent people, organization, countries, but also computers, species, molecules or concepts), and (2) links (which may represent any recognizable connection between nodes like friendship among people, collaboration among organizations, geographic neighbourhood between countries, a wireless computer network, food chains in an ecosystem, connections between molecules or the essencal semantic connection between concepts in a language). If the links are directed (e.g. communication with messages, the spreading of a virus, influence of power etc.) then the network is directed. If the connection is measurable, then the network is weighted.

Formally, networks are represented in form of mathematical graph structures which are ordered pairs $G = (N, E)$, whereby $N = \{ n_1, n_2, ..., n_m \}$ is the set of nodes, and $E = \{ (n_i, n_j) \mid n_i, n_j \in N \}$ a set of edges or links. If the pairs in E are ordered, we call G a directed graph or digraph.

Networks are often due to simplicity of computation, represented in the form of the adjecency matrix $A = [a_{ij} \mid a_{ij} \in \{0,1\}]$ which is of size m × m where m is the number of nodes in the network. The elements equal to 1 if there exists a connection between the corresponding two nodes, 0 otherwise. If the network is undirected, the matrix is symetric. If the network is weighted, instead of 1, one can write the actual weight of a given connection.

For the current research the concepts of bipartite, tripartite and later on n-partite graph are of special importance. A bipartite graph is a special graph $G = (N_1 \cup N_2, E)$ in which there exists a partition of the node set such that if an edge has one end in N_1 then the other end of the edge is in N_2, which means that there are no connections between nodes inside any of the sets in the partition, only between them. In a tripartite graph there are three such sets, while in a n-partite graph there are n-such sets.

An example of a bipartite graph is a network of authors (A) and publications (P), in which nodes are autors and publications (there is a partition of the node set into two distinct sets), while the connections are the essential authorship relation. As one can see, there will never be a connection between authors (e.g. authors do not write other autors), nor among publications (e.g. publications are not written by other publications). An example of a tripartite graph can be the network of authors (A), publications (P) and keywords used on publications (K). In the following we shall analyze the 4-partite graph of authors (A), publications (P),keywords (K) and scientific fields (F).

Another characteristic of n-partite graphs that needs to be pointed out here is the possibility to represent it by n (n −1)-partite graphs (e.g. a 4-partite graph can be represented by 4 tripartite graphs, a tripartite by 3 bipartite graphs etc.). Lower level graphs are constructed by ommitting all nodes from a given partition together with all edges the nodes participate in. The bipartite graph representation is more practical then the 4-partite and tripartite since bipartite graphs can be represented in simple matrix form. For example the graph AP can be represented in a matrix in which columns are authors (elements from A) and rows are publications (elements from P) while we put a 1 in each element of the matrix if the corresponding author wrote the corresponding publication, or 0 otherwise.

Graph folding [16] is a mapping that maps from one graph to another by always je mapping nodes into nodes and edges into edges. For our investigation it is sufficient to state that by a special kind of graph folding we can acquire unipartite from bipartite graphs by matrix multiplication with the transposed matrix. Let |AK| be the matrix of authors and keywords used by them on some publications, then by folding the graph with the operation $|AK||AK|^T$ we acquire a matrix that represents the social network of authors which used the same keyword, e.g. two authors will be connected if they used the same keyword on some of their publications. The dual matrix |AK|T|AK| is the conceptual network of keywords in which to keywords are connected if they were used by the same author.

We will use this procedure quite intensively for the construction of a number of conceptual and social networks. Mika [9] used a very similar procedure with tripartite graphs to define a folksonomy (the ACI model) of the social tagging application Delicious[2]. Hereafter we will construct the following graphs:

- $|AK|^T|AK|$ – the conceptual network of keywords for each scientific field (biomedical, biotechnical, social, natural, technical sciences and the humanities),
- $|AF|^T|AF|$ – social network of scientific fields according to the collaboration of scientists (e.g. two fields are connected if a scientist published in both fields),
- $|KF|^T|KF|$ – social network of scientific fields according to conceptual connections (e.g. two fields are connected if the same keyword has been published in both of them).

Since some of the analysed conceptual networks are complex, an adequate visualization algorithm is needed. Herein we decided to use the k-core decomposition algorithm described in [17]. The description of this algorithm goes beyond the objectives of this article and has thus been omitted. For our purposes it is enough to state that the algorithm finds cores that represent mutually well connected nodes.

DATA HARVESTING AND IMPLEMENTATION

Since all data about CROSBI was available on the Web it was neccessary to implement an adequate spider program that would harvest the data. To achieve that the programming language Python[3] and specifically the module Scrapy[4] was used which tremendeously simplified the implementation. To extract the semistructured data xPath and regular expressions were used. For each publication the following data was gathered:

- author's name and surname,
- year of publication,
- scientific fields of publication,
- keywords,
- CROSBI key (for unique identification and later analysis),
- type of publication.

The harvesting was conducted in November 2010 and data about a total of 285 234 scientific publications were collected. Since CROSBI is a social application where authors by themselves entered the data, there was need for clensing (wrong syntax, duplicates, special characters, wrong encoding etc.). The cleaned data was stored in a PostgreSQL database[5] for later analysis.

By using a number of queries conceptual networks for each scientific field was constructed for the period 2000-2010. Due to a large number of keywords, combinatoric explosion and limited hardware capabilities only those keywords which were used more than 200 times were used for further computation (there was a total of 368 378 keywords, while 651 were used more than 200 times). The constructed networks were visualized using the LaNet-vi[6] tool.

To categorize the keywords identified in the cores of each conceptual network for each scientific field, the Wikipedia[7] application programming interface was used, especially its coratian[8], english[9] and german[10] version. Wikipedia allows its users in addition to hypertextual data to enter metadata about each term like categories which apply to a given term. These meta-data are shown at the top of each page (standard categories like "Article which need additional citations" or "Articles which should be merged") as well as on the bottom of it (user-defined categories). For the purpose of categorizing keywords across publications only user-defined categories were used since these, as opposed to standard categories, reflect certain semantics provided by the social system of users. In accordance with this again Python was used to collect the adequate categories for identified keywords, and XSB Prolog[11] was later used to implement a logic program that connected similar concepts into clusters. The logic program consisted of a number of simple rules. One of the most important rules was that two concepts will fall into the same cluster if they have at least one common category according to the categorization of Wikipedia users.

By using additional queries the social networks of scientific fields based on scientific collaboration and conceptual connection were constructed and stored in a ZODB object-base[12] for later processing and analysis. The Python module[13] NetworkX was used to visualize the networks.

ANALYSIS

The analysis in divided into two parts: firstly we analyse the conceptual networks of particular scientific fields ($|AK|^T|AK|$ graphs), and secondly the social networks based on scientific collaboration (graph $|AF|^T|AF|$) and conceptual connection (graph $|KF|^T|KF|$).

CONCEPTUAL NETWORK ANALYSIS

Figure **1** shows the conceptual graph of biomedical sciences[14]. The size of the nodes denotes the number of connections the node participates in, while the color of the node simbolizes the number of connections of the node in the given graph. The nodes shown in the middle of the graph represent the best connected nodes, mutually as well as to other nodes in the network. One could state that these are the core concepts of croatian publications in biomedical research of the observed period.

Of special importance is the topology of the conceptual network. Since we can observe one main, one secondary (in the upper left of the middle) and one marginal (in the lower left) core, we can conclude that biomedical research is quite well focuse on a small number of well-connected research areas.

The conceptual core of biomedical research contians 76 concepts, 44 (57,89 %) in the English language and 32 (42,11 %) in Croatian. By using the previously mentioned logic program based on the analysis of Wikipedia categories, conceptual clusters were obtained and are summarized in Table **1**. Uncategorized concepts are concpts for which there is no applicable

Table 1. Conceptual clusters of biomedical sciences.

Category	Keywords
Aging Gerontology	aging, elderly
Aging-associated diseases	hypertension, stroke
Cardiology	atherosclerosis, hypertension
Diabetes	diabetes, diabetes mellitus
Epidemiology	epidemiology, prevalence, risk factors
Gender	gender, women
Greek loanwords	diagnosis, epilepsy, genetics, pain
Medical conditions related to obesity	diabetes, diabetes mellitus, hypertension
Medical statistics	prevalence, risk factors
Medical terms	diagnosis, epilepsy, prognosis
Nutrition	cholesterol, diabetes, diabetes mellitus, obesity
Sociology	elderly, gender
Antropologija	*dijete, djeca*
Bolesti živčanoga sustava	*epilepsija, moždani udar*
Medicina	*dijagnostika, dijagnoza, terapija, zdravlje*
Medicinska dijagnostika	*dijagnostika, dijagnoza*
U izradi, Bolesti i poremećaji	*epilepsija, moždani udar*
U izradi, Medicina	*terapija, zdravlje*
Čovjek	*dijete, djeca, žene*
Uncategorized concepts	*ateroskleroza, edukacija, etiologija, liječenje, prevencija, rehabilitacija, starenje, starije osobe*
Unaligned concepts	Croatia, Europe, *Hrvatska*, PTSD, PTSP, Zagreb, apoptosis, breast cancer, children, *depresija*, Doppler, depression, education, *epidemiologija, etika*, gender differences, *genetika, hipertenzija*, incidence, knowledge, *kvaliteta života*, management, mortality, multiple sclerosis, oxidative stress, pregnancy, prevention, *pušenje*, quality of life, rat, screening, stres, stress, therapy, treatment, *trudnoća*, ultrasound, *ultrazvuk*, war, *šećerna bolest*

category on Wikipedia, while unaligned concepts are thode for which there is at least one category, but there are no other concepts that have any mutual category and thus those concepts cannot be alignet into any cluster. These two groups of concepts are of special importance for our investigation since they indicate misalignement between the two observed social systems. We should also state here that categories denoted in Croatian as *U izradi* (eng. Under construction), are categories which are not yet fully finished on Wikipedia, which means that not all relevant concepts have been categorized into them.

From Table 1 one can read that the croatian biomedical sciences in the observed period mostly dealt with aging, different types of diseases and epidemics, the relationship between people and health condition, nutrition, and generally medicine and connected terms.

According to the interpretation, we can observe that the public as a social system perceives a connection between this field and sociology, anthropology and the usage of Greek terms.
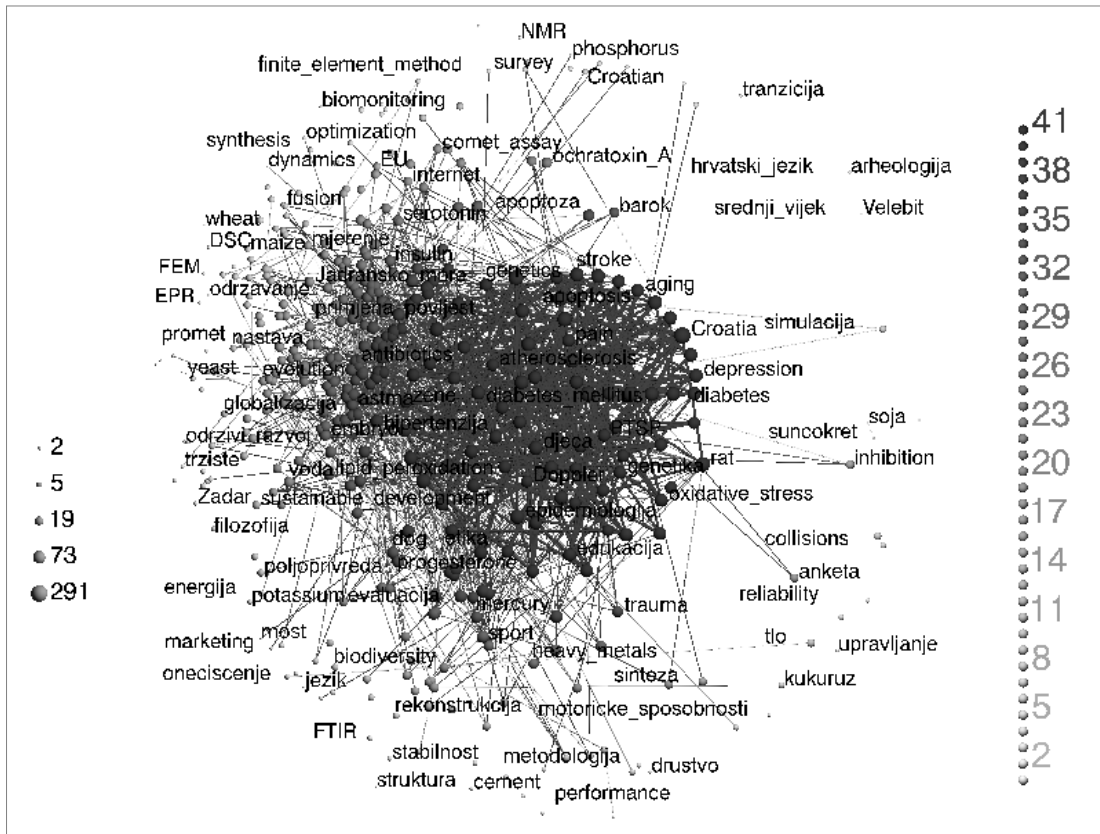
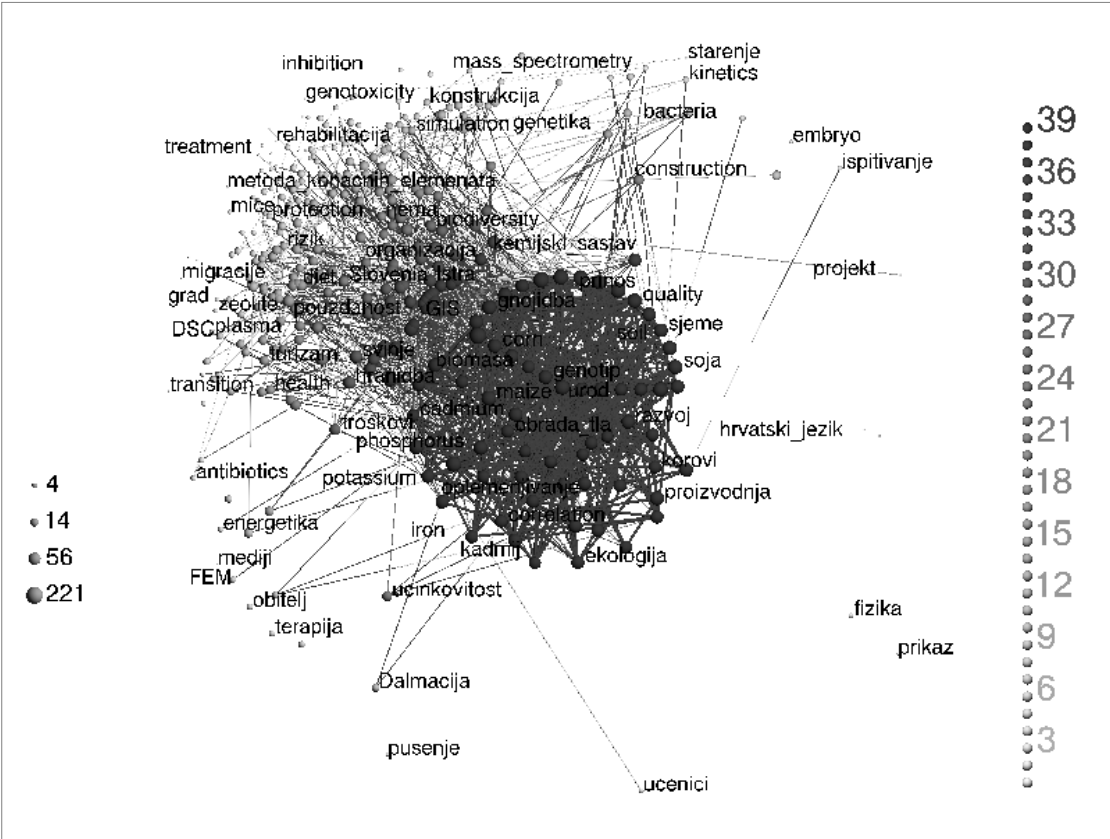**Figure 1.** Conceptual network of croatian biomedical sciences 2000-2010.

Figure 2. Conceptual network of croatian biotechnical sciences 2000-2010.

Table 2. Conceptual clusters of biotechnical sciences.

Categories	Keywords
Agriculture in Mesoamerica, Crops originating from Mexico, Crops originating from the Americas, Demulcents, Flora of Guatemala, Flora of Mexico, Fruit vegetables, Grasses of Mexico, Maize, Mexican ingredients, Model organisms, Native American cuisine, Plants described in 1753, Tropical agriculture, Zea	corn, maize
Biology and pharmacology of chemical elements	cadmium, iron, nitrogen, phosphorus, potassium, zinc
Chemical elements	cadmium, calcium, iron, nitrogen, phosphorus, potassium, zinc
Dietary minerals	calcium, iron, phosphorus, potassium, zinc
Energy crops	corn, maize, soybean, wheat
Nonmetals, Pnictogens	nitrogen, phosphorus
Pyrotechnic fuels	iron, phosphorus, zinc
Reducing agents	calcium, potassium, zinc
Sequenced genomes	corn, maize, soybean
Staple foods	corn, maize, wheat
Transition metals	cadmium, iron, zinc
Wheat	wheat, winter wheat
Fizikalne veličine	*energija, temperatura*
Industrijske biljke	*soja, šećerna repa*
Kemijski elementi U izradi, Kemijski elementi	*dušik, kadmij*
Poljoprivreda	*navodnjavanje, poljoprivreda*
Trave, Žitarice	*kukuruz, pšenica*
U izradi, Botanika	*kukuruz, pšenica, šećerna repa*
Uncategorized concepts	*bolesti, genotip, gnojidba*, grain yield, *hibridi, kakvoća, kultivar, obrada tla, oplemenjivanje, ozima pšenica, prinos, prinos zrna, razvoj, sorta, sorte, stabilnost, teški metali, urod, urod zrna*
Unaligned concepts	Croatia, *Hrvatska, biomasa*, correlation, *ekologija*, fertilization, *korovi, kvaliteta*, precipitation, pH, *proizvodnja*, protein, quality, resistance, *sjeme*, soil, stability, *suncokret*, temperature, *tlo, vinova loza*, yield

Figure 2 shows the visualization of the conceptual network of the biotechnical sciences in the observed period. As we can see from the analysis of the topology of the network, there is one main (middle of figure), and one secondary (upper left of the middle) core. This indicates

even greater focus of researc if compared to the biomedical sciences. The conceptual core consists of 63 keywords, 25 (39,68 %) in English and 38 (60,32 %) in Croatian.

The public as a social system can best interpret concepts from the biotechnical sciences regarding various agricultures (corn, maize, soybean, wheat, and cro. *šećerna repa* – sugar beet), as well as various chemical elements. Another important interpreted subject is agriculture (cro. *poljoprivreda*). There are indications that a connection to physics and chemistry is anticipated.

Figure 3 shows the conceptual network of the social sciences. Besides the main core, we can observe at least two secondary and a number of marginal cores. These scattered concepts indicate less focus in research as well as a broadness of the investigated subjects in the social sciences. The main core contains 73 concepts, 21 (28,77 %) in English and 52 (71,23 %) in Croatian.

By using the logic program the clustering as in Table 3 is obtained. As can be seen from Table 3, from the perspective of the public, the social sciences mostly investigated the relationship between Croatia and the European Union, a number of concepts bound to economy, culture, nurture and education, politics, as well as tertiary activities. It is interesting to observe that this field of research is also connected with the very concept of science itself as well as with the human which is at its centre. The public also anticipates the connection of the social sciences with ecology and applied sciences.

Figure 4 shows the conceptual network of the croatian humanities. Except for a maim core, we can observe an outstanding secondary as well as a number of marginal which are all well connected. Thus the humanities seem to be quite focused with a number of well connected boundary areas. The core of the humanities includes 52 keywords, with only two in English (3,85 %), 49 (94,23 %) in Croatian and one (1,92 %) in German. This focus on croatian publishing does not come as a surprise, since some of the main concerns of these sciences exclusively deal with Croatia and its heritage.

The clusters obtained through the logic program are shown in Table 4.

The humanities, as from the perspective of the public, mostly dealt with Slavonia and Dalmatia (two great regions of Croatia), philsophy (cro. *filozofija*), culture (cro. *kultura*), history (cro. *povijest*), heritage (cro. *baština*), literature (cro. *književnost*) and arts (cro. *umjetnost*). The public seems to anticipate a connection between the humanities and the social sciences as well as geography. Especially interesting is the wrong interpretation (at least what croatian scientific classification concerns) that archaeology is part of the social sciences. Also, as with the social sciences, the humanities are being connected to the very concept of science.

Figure 5 shows the conceptual network of the natural sciences of Croatia. As opposed to other conceptual networks, besides the main core, we see quite a number of secondary and marginal cores, which are visually hardly to differentiate. This situation is expected since the natural sciences consist of a number of essentially quite different sciences. The main core consists of a total of 52 keywords, 30 (57,69 %) in English and 22 (42,31 %) in Croatian.

Table 5 provides the results of applying the logic program to the natural sciences core concepts.

When it comes to the natural sciences, the public seems to mostly anticipate concepts from biology and geography as well as their connected areas. There is an indication that the natural sciences might be connected to (geo)politics (e.g. categories European countries, Member states of the Union for the Mediterranean, Liberal democracies, Members of the North Atlantic Treaty Organization) and ecology.
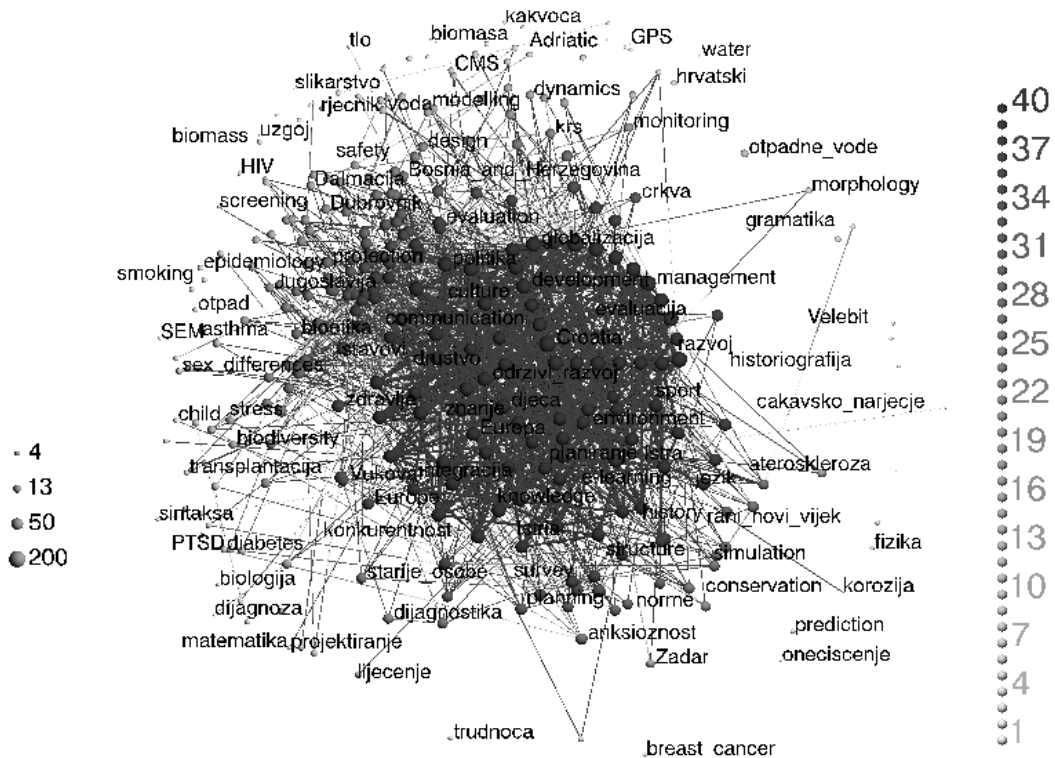
Figure 3. Conceptual network of croatian social sciences 2000-2010.

Table 3. Conceptual clusters of social sciences.

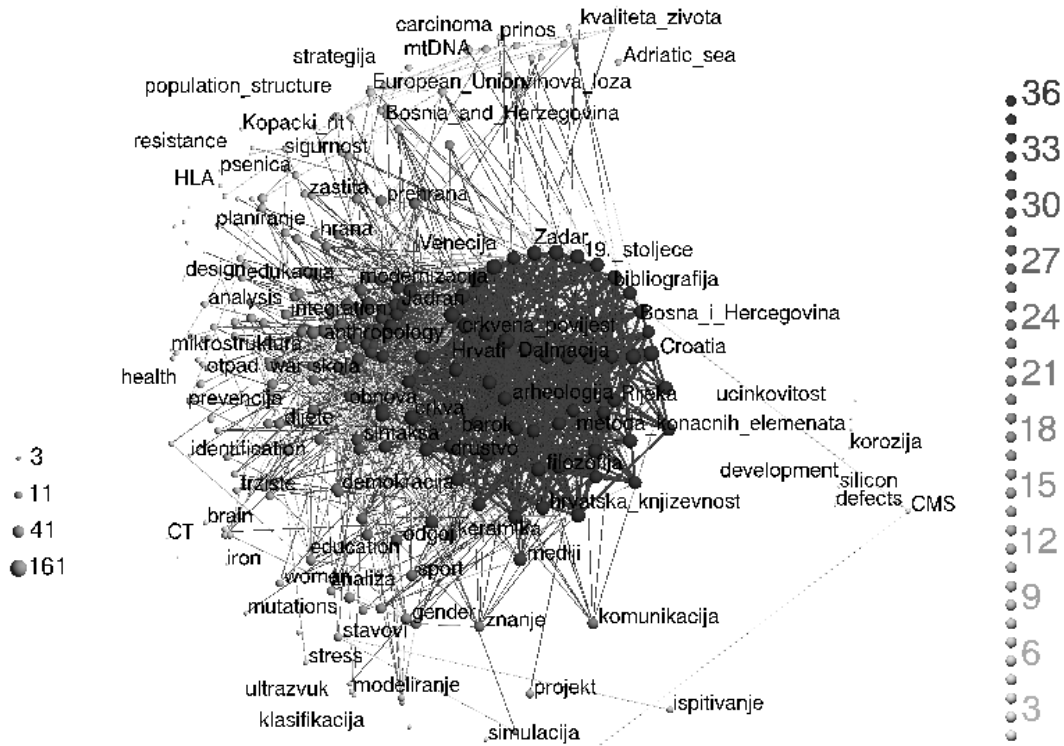| Categories	Keywords
1993 establishments, European Union, Federalism, International organizations of Europe, Political system	EU, European Union
Države u Europi, Hrvatska	*Hrvatska, Republika Hrvatska*
Ekologija	*ekologija, zaštita okoliša*
Ekonomski pojmovi	*globalizacija, gospodarstvo, tranzicija, tržište*
Kultura	*kultura, mediji, turizam*
Menadžment	*kvaliteta, menadžment, poduzetništvo*
Odgoj i obrazovanje	*obrazovanje, odgoj, škola*
Politika	*demokracija, globalizacija, politika, tranzicija*
Primijenjene znanosti	*komunikacija, mediji, promet, tehnologija*
Tercijarne djelatnosti	*promet, turizam*
Znanost	*istraživanje, metodologija, znanost*
Čovjek	*djeca, žene*
Uncategorized concepts	*konkurentnost, mladi, model, modernizacija, održivi razvoj, razvoj, studenti, upravljanje, učinkovitost, zaštita*
Unaligned concepts	Croatia, *Europa, Europska unija*, Internet, Istria, communication, competitiveness, culture, development, *društvo*, education, *etika*, globalization, *identitet*, integration, knowledge, management, *marketing, obitelj, okoliš, organizacija, planiranje, povijest*, quality, *religija*, research, *sigurnost, strategija*, sustainable development, *šport*, technology, tourism, transition, *znanje*

Figure 4. Conceptual network of croatian humanities 2000-2010.

Table 4. Conceptual clusters of humanities.

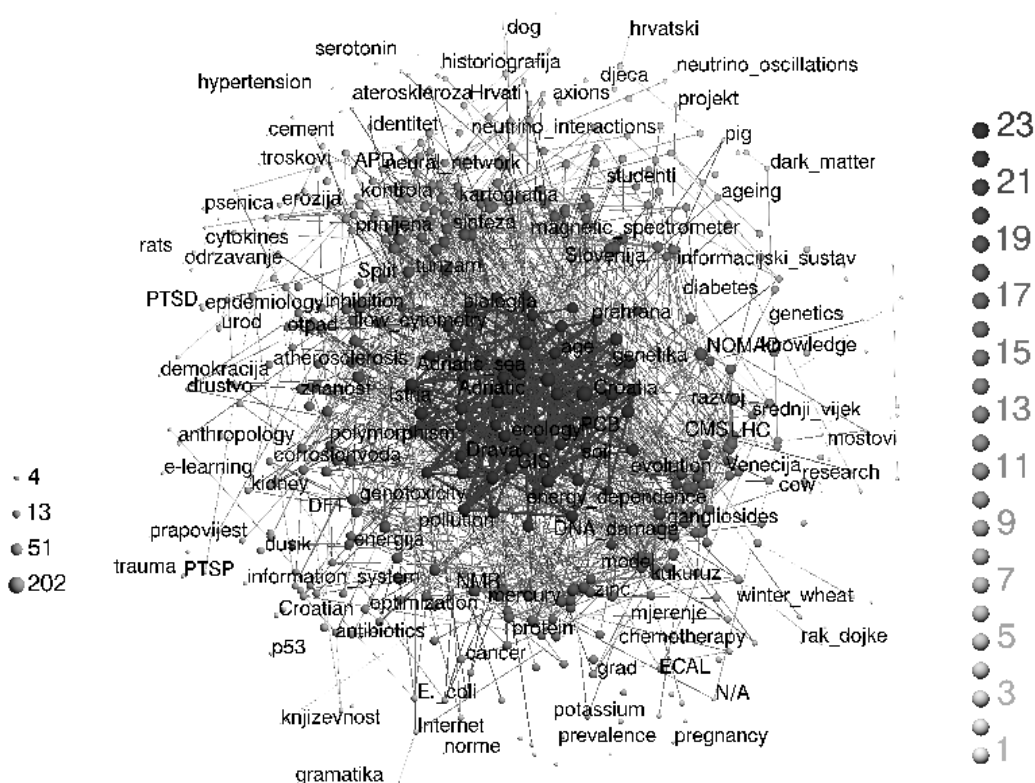| Categories	Keywords
Baranjski leksikon, Osijek	Osijek, *Slavonija*
Društvene znanosti	*arheologija, politika*
Filozofska terminologija	*politika, umjetnost*
Kultura	*književnost, kultura*
Povijesna znanost	*historiografija, povijest, rat*
Povijest književnosti, Renesansa i novi vijek, Umjetnička razdoblja	*barok, renesansa*
Svjetska baština u Hrvatskoj	Dubrovnik, Split
Zemljopis Hrvatske	*Dalmacija, Slavonija*
Znanost	*istraživanje, metodologija, znanost*
Uncategorized concepts	*crkvena povijest, migracije, prikaz, rani novi vijek, rani srednji vijek*
Unaligned concepts	19. *stoljeće, Bosna i Hercegovina*, Croatia, Dalmatia, *Europa, Hrvati, Hrvatska, Istra, Jugoslavija*, Kroatien, *Rijeka, Venecija*, Zadar, Zagreb, *arhitektura, bibliografija, biografija, crkva, društvo, filozofija, gospodarstvo, grad, hrvatska književnost, hrvatski jezik, identitet, jezik, religija, slikarstvo, srednji vijek*

Figure 5. Conceptual network of croatian natural sciences 2000-2010.

Table 5. Conceptual clusters of natural sciences.

| Categories	Keywords
Adriatic Sea	Adriatic, Adriatic sea
Aquatic ecology	fish, phytoplankton
Countries of the Mediterranean Sea, European countries, Member states of the Union for the Mediterranean, Slavic countries	Bosnia and Herzegovina, Croatia, Slovenia
European seas	Adriatic, Adriatic sea, Mediterranean
Liberal democracies, Members of the North Atlantic Treaty Organization, States and territories established in 1991, Serbo-Croatian-speaking countries	Croatia, Slovenia Bosnia and Herzegovina, Croatia
Jadransko more, *Mora*	*Jadran, Jadransko more*
Planine Hrvatske	Medvednica, Velebit
Zemljopis Hrvatske	Dalmacija, *krš*
Uncategorized concepts	*rast, zaštita, zooplankton*
Unaligned concepts	*Bosna i Hercegovina*, Dalmatia, Dinarides, Drava, Dubrovnik, Europe, GIS, *Hrvatska, Istra*, Istria, PCB, Zagreb, age, biodiversity, *biologija*, conservation, distribution, ecology, *ekologija*, environment, fauna, *fitoplankton*, flora, geochemistry, growth, karst, monitoring, *morfologija*, morphology, mortality, pollution, *sediment*, soil, temperature, *tlo*
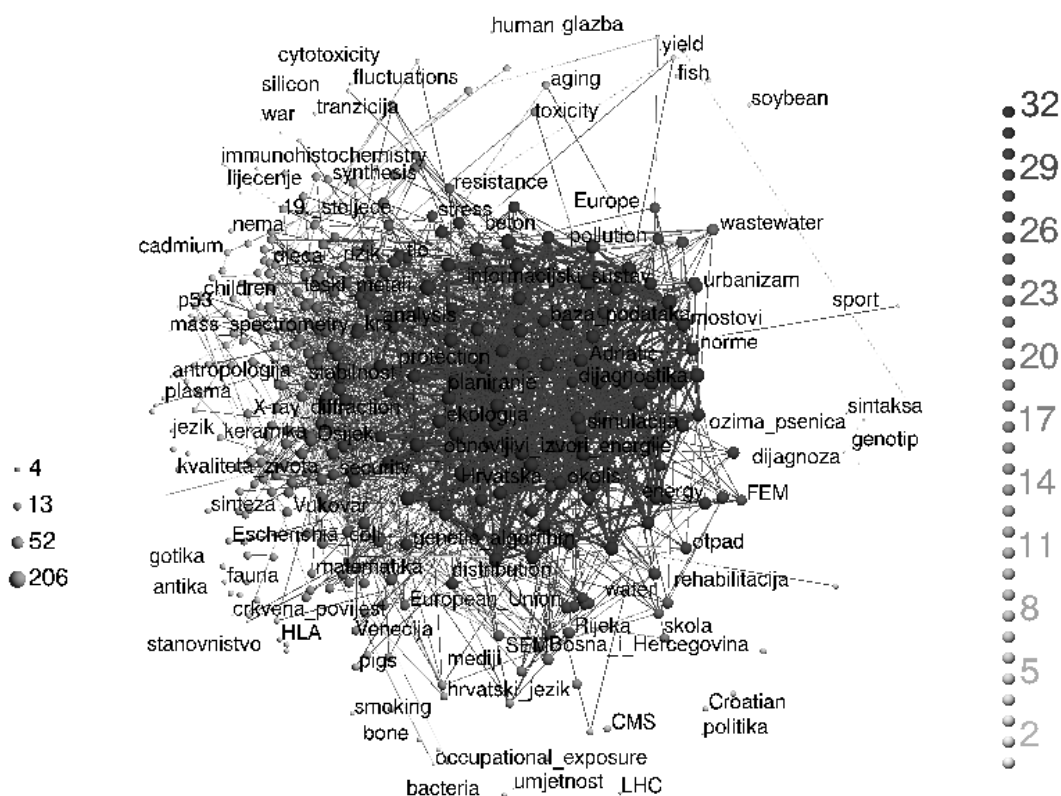
Figure 6. Conceptual network of croatian technical sciences 2000-2010.

In the end of the conceptual analysis in Figure 6 the conceptual network of the technical sciences are shown. In this case we have a main, one distinct secondary and four marginal cores. From these results one may conclude that the technical sciences were less focused and investigated a broad spectrum of various fields. The main core consists of 86 concepts, 36 (41,86 %) in English and 50 (58,14 %) in Croatian.

The logic program resulted with the clusters of concepts as in Table 6.

Based on Table 6 one can deduce that the technical sciences are the least understood by the public as a social system. The best understood concepts deal with construction engineering and computer science (in which we can include the category American inventions). The public connects the technical sciences more than all other fields with various concepts regarding science and technology (research methods, thought, as well as cro. *tehnologija i znanost* – technology and science). The public also connects these core concepts with various fields like control (cro. *upravljanje*), operations research (cro. *operacijska istraživanja*), ecology (cro. *ekologija*), physics (cro. *fizika*), chemistry (cro. *kemija*), medicine (cro. *medicina*) and applied sciences (cro. *primjenjene znanosti*). An interesting contextual semantic error is also that the word *modelling* (British English) and *modeling* (American English) are categorized into surnames.

SOCIAL NETWORK ANALYSIS

Figure 7 depicts the social network of the scientific fields according to the criteria of scientific collaboration. The size of the nodes implies the total number of publications in the given field, while the wideness of the link implies the number of authors which published in both connected fields. The annotated connection weights are the percentage of the total number of connections in the graph whereby all recursive connections (connections from the node to itself) are omitted.

Table 6. Conceptual clusters of technical sciences.

Categories	Keywords
American inventions	GPS, Internet
Computational neuroscience, Data mining, Econometrics, Network architecture, Networks, Neural networks, Information, knowledge and uncertainty	neural network, neural networks
Management	management, planning
Operations research	mathematical model, simulation
Research methods	analysis, research
Surnames	modeling, modelling
Thought	analysis, planning
Ekologija	*ekologija, zaštita okoliša*
Fizikalne veličine	*energija, temperatura*
Građevinarstvo, Mostovi, U izradi, Promet	*most, mostovi*
Građevinski materijali	*beton, cement*
Kemija	*korozija, zaštita okoliša*
Medicina	*dijagnostika, zaštita okoliša*
Menadžment	*kvaliteta, proračun*
Primijenjene znanosti	*promet, tehnologija*
Tehnologija	*struktura, tehnologija*
Znanost	*istraživanje, mjerenje*
Uncategorized concepts	*konstrukcija*, mechanical properties, *mehanička svojstva, metoda konačnih elemenata, model, modeliranje, norme, obnova, održivi razvoj, pouzdanost, primjena, projektiranje, razvoj, sanacija, simulacija, trajnost, troškovi, upravljanje, zaštita*
Unaligned concepts	Adriatic, Croatia, GIS, *Hrvatska*, Zagreb, *analiza, baza podataka*, concrete, construction, corrosion, design, development, education, environment, finite element method, maintenance, measurement, monitoring, *obrazovanje, održavanje, okoliš, optimizacija*, optimization, *planiranje*, povijest, prediction, *proizvodnja, projekt*, protection, quality, reliability, safety, *sigurnost*, structure, sustainable development, temperature, *urbanizam, voda*

From Figure 7 one can see that the natural sciences are, as expected, most well connected, while the humanities are the least well connected with the other fields. Additionally the strongest connection is between the natural and the technical sciences, then among the natural and the biotechnical and then the natural and biomedical sciences. The weakest connection is between the humanities and the biotechnical sciences.

Figure 8 shows the social network between scientific fields according to conceptual similarity. The size of the nodes again depicts the total number of publications in the given field, whilst the width of the links shows the number of shared keywords. Even if here the natural sciences are again most well connected, there is only a slight difference to the social

sciences (0.05 %). Again, the least well connected are the humanities. This time the strongest
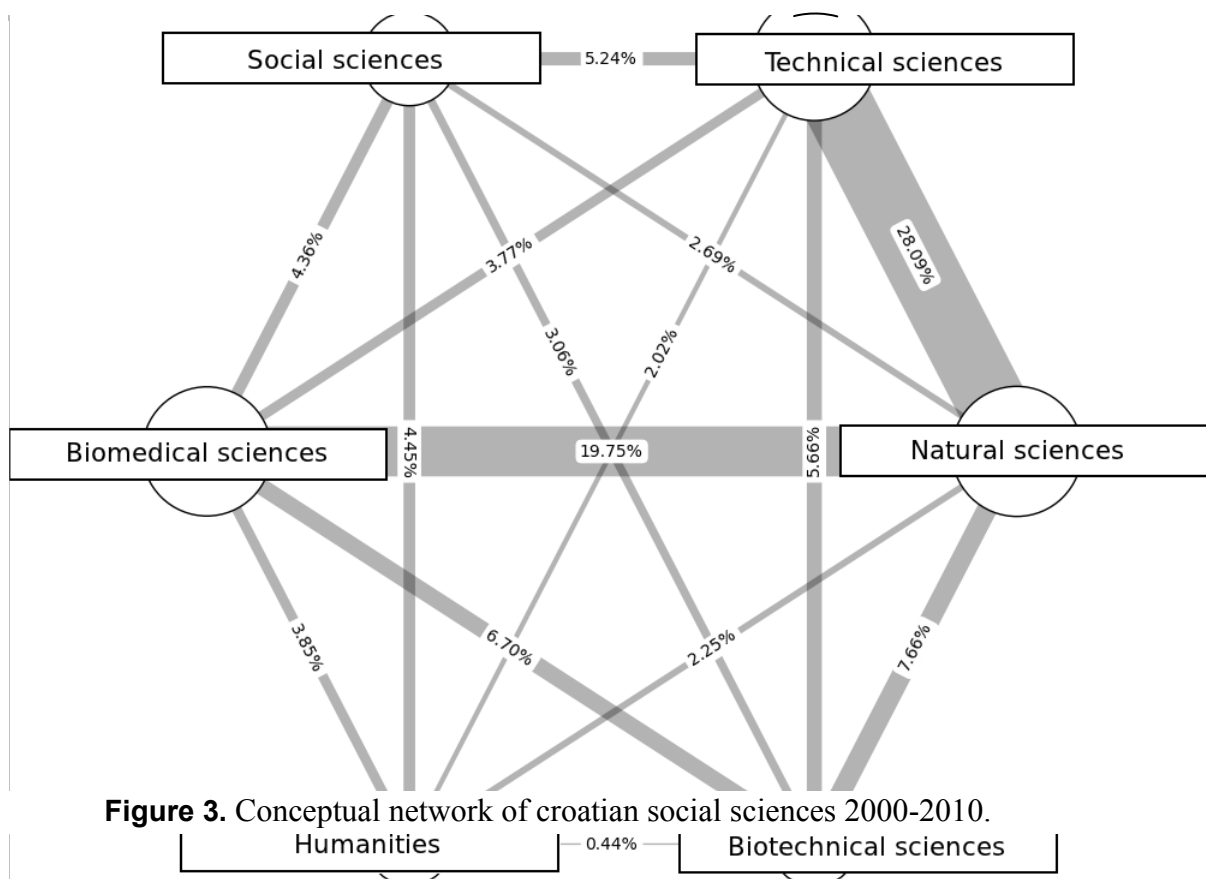

Figure 3. Conceptual network of croatian social sciences 2000-2010.

Figure 7. Connections among scientific fields in Croatia (social criteria) 2000-2010.

linkage is between the natural and the social sciences, then between the natural and the biomedical sciences, and then between the social and the biomedical sciences. The weakest connection is again between the humanities and the biotechnical sciences.

DISCUSSION

The uncategorized and unaligned concepts from the conceptual analyses need special attention. Table **7** summarizes the total number of such concepts. The uncategorized concepts indicate that the public as a social system do not anticipate a category in which the concept might be classified, e.g. the system is not able to interpret the given term. The unaligned concepts on the other hand, indicate that even if the social system is able to interpret these terms, it does not connect them with any other concept from the main core of a given scientific field. This inability of connecting is disturbing, since these are the most well connected concepts of each field.

From Table 7 we can conclude that the number of concepts the social system interprets only partially or not at all is quite high (on average over 60 %) which confirms hypothesis H1, e.g. there is a disproportion between the subjects the croatian scientific community investigates and the understanding of these subjects in the public. In order to address this problem there is need to intensify the transfer of knowledge from the scientific community to the public.

From the social network analyses from both the social and the conceptual perspective we can conclude about the disproportion between the conceptual interconnection of scientific fields and the collaboration between the fields. Table 8 gives a detailed overview of this disproportion and is calculated with the absolute difference between the adjacency matrices of both graphs.
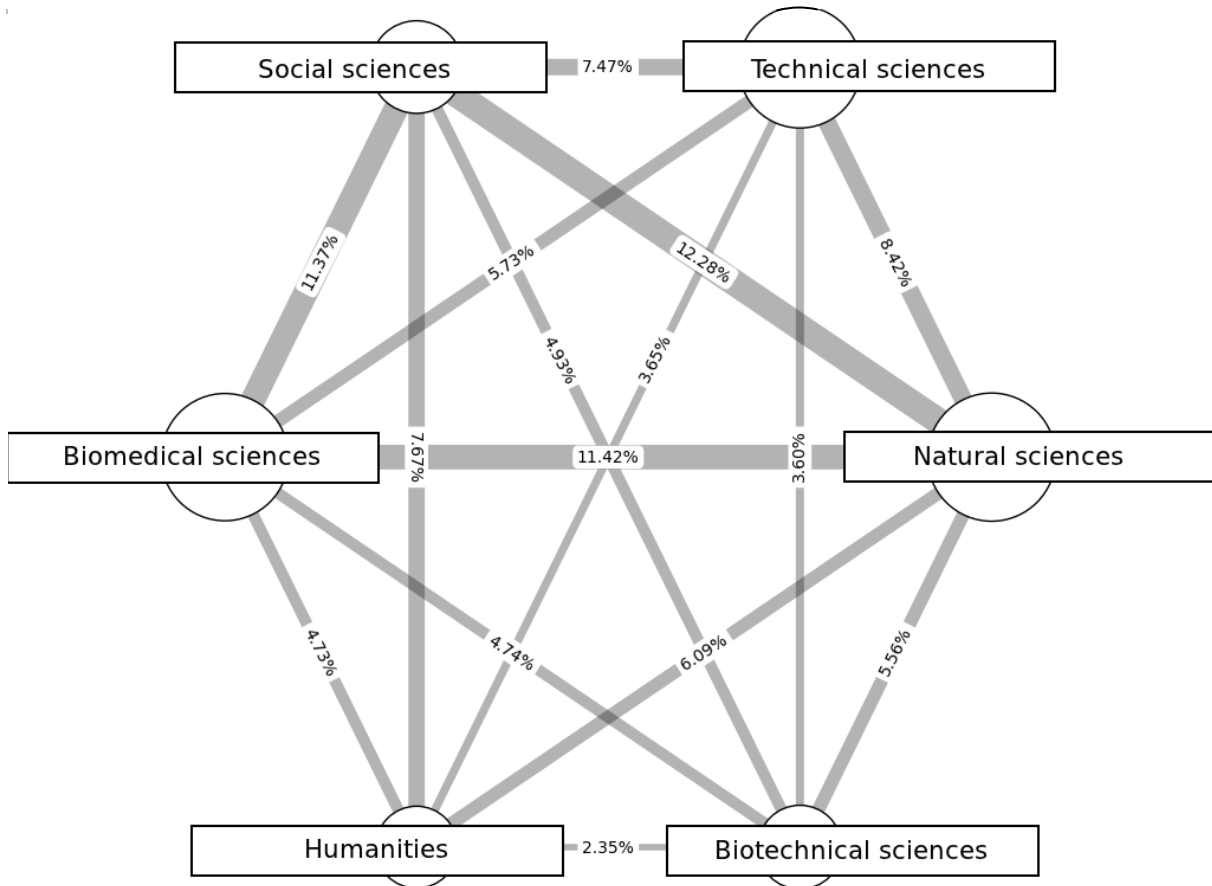
Figure 8. Connections among scientific fields in Croatia (conceptual criteria) 2000-2010.

Table 7. Uncategorized and unaligned concepts.

Concepts	Biomedical sciences	Biotechnical sciences	Social sciences	Humanities	Natural sciences	Technical sciences
Uncategorized concepts	8	19	10	5	3	19
Unaligned concepts	40	22	35	29	35	38
Total	48	41	45	34	38	57
Total, %	63,16	65,08	61,64	65,38	73,08	66,28

Table 8. Disproportion of social networks according to social and conceptual criteria.

Scientific areas	Biomedical sciences	Biotechnical sciences	Social sciences	Humanities	Natural sciences	
Biotechnical sciences	3,98 %					
Social sciences	**10,87** %	4,58 %				
Humanities	4,29 %	2,30 %	7,17 %			
Natural sciences	**9,17** %	4,69 %	**11,97** %	5,83 %		
Technical sciences	5,30 %	2,96 %	6,87 %	3,42 %	5,22 %	
Total						**88,61 %**

The total disproportion between these two networks is rather high, 88,61 % which confirms H2. This disproportion is in particular obvious between social, natural and biomedical sciences. In accordance with these results interdisciplinary research in the identified areas should be stimulated and encouraged.

CONCLUSIONS

This article provided a first conceptual and social network analysis of the croatian scientific bibliography. The obtained results were interpreted through the reflection (Wikipedia) of the public as a social system. By using a number of contemporary methods from network theory the conceptual networks of each scientific field were visualized which allowed us to reason about the focus and broadness of research. The field of the natural sciences should be analyzed in more detail, since it consists of a number, essentially quite different sciences. Such an analysis would provide better insight into the main concepts of the field.

By using a logic program the obtained core concepts of each field were interpreted by using user-defined categories from Wikipedia. This allowed us to gather all concepts that the public as a social system was not able to interpret adequately. The number of such concepts is quite high (on average over 60 %) which indicates a disproportion of the croatian scientific community and the public. It is thus necessary to stimulate and intensify the transfer of knowledge to the public through popular scientific publications and happenings, the social Web as well as more media presence of scientific themes.

Additionally an analysis of the disproportion between the social networks of scientific fields according to social and conceptual criteria was provided. The results show that there is a clear and surprisingly high disproportion between conceptually connected fields of research and actual interdisciplinary research. These results indicate the necessity of stimulation and encouragement of interdisciplinary research, especially between the biomedical, social and natural sciences.

REMARKS

[1] http://bib.irb.hr,
[2] http://www.delicious.com,
[3] http://www.python.org,
[4] http://scrapy.org,
[5] http://www.postgresql.org,
[6] http://xavier.informatics.indiana.edu/lanet-vi,
[7] http://www.wikipedia.org,
[8] http://hr.wikipedia.org/w/api.php,
[9] http://en.wikipedia.org/w/api.php,
[10] http://de.wikipedia.org/w/api.php,
[11] http://xsb.sourceforge.net,
[12] http://www.zodb.org,
[13] http://networkx.lanl.gov,
[14] All graphs in this article are given in low resolution and gray-scale. High resolution coloured graphs are available at http://autopoiesis.foi.hr/bib.

REFERENCES

[1] Schatten, M.: *Autopoietic Theory as a Framework for Biometrics.*
Interdisciplinary Description of Complex Systems **6**(2), 74-79, 2008,
http://indecs.eu/2008/indecs2008-pp74-79.pdf,

[2] Schatten, M.; Bača, M. and Ivanković, M.: *Public Interfaces as the Result of Social Systems Structural Coupling.*
In Mertik, M. and Damij, N., eds.: 1ˢᵗ International Conference on Information Society and Information Technologies. Proceedings, Faculty of information studies, Novo Mesto, 2009, http://bib.irb.hr/datoteka/430152.CS_03.pdf,

[3] Luhmann, N.: *Soziale Systeme: Grundriß einer allgemeinen Theorie*.
Suhrkamp, Frankfurt, 1984,

[4] Schatten, M. and Bača, M.: *A Critical Review of Autopoietic Theory and its Applications to Living, Social, Organizational and Information Systems*.
Društvena Istraživanja **108-109**(4-5), 837-852, 2010, http://hrcak.srce.hr/file/90557,

[5] Frawley, W.J.; Piatetsky-Shapiro, G. and Matheus, C.J.: *Knowledge Discovery in Databases: An Overview.*
AI Magazine, 57-70, 1992, http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1992.pdf,

[6] Berners-Lee, T.; Hendler, J. and Lassila, O.: *The Semantic Web.*
Scientific American Magazine, 2001,

[7] Mika, P.: *Flink: Semantic Web technology for the extraction and analysis of social networks.*
Web Semantics: Science, Services and Agents on the World Wide Web **3**(2-3), 211-223, 2005, http://www.cs.vu.nl/~pmika/research/papers/JWS-Flink.pdf,

[8] Mika, P.: *Social Networks and the Semantic Web.*
Springer, New York, 2007b,

[9] Mika, P.: *Ontologies are us: A unified model of social networks and semantics.*
Web Semantics: Science, Services and Agents on the World Wide Web **5**(1), 5-15, 2007a, http://www.cs.vu.nl/~pmika/research/papers/ISWC-folksonomy.pdf,

[10] Borić, V.: *Citation Analysis of the Papers' Published in Acta Stomatologica Croatica using Web of Science Database.*
Acta stomatologica Croatica **42**(2), 123-139, 2008, http://hrcak.srce.hr/file/40800,

[11] Borić, V. and Strujić, M.: *Bibliometric Analysis of Acta Stomatologica Croatica:* 1987-2006.
Acta stomatologica Croatica **40**(4), 336-346, 2006, http://hrcak.srce.hr/file/17935,

[12] Vilibić, I.: *Bibliometric analysis of the Adriatic-related oceanography and meteorology publications*.
Geofizika **26**(2), 229-243, 2009, http://geofizika-journal.gfz.hr/vol_26/No2/26_12_Vilibic.pdf.

[13] Aparac, T. and Pehar, F.: *Information Sciences in Croatia: A View from the Perspective of Bibliometric Analysis of two Leading Journals.*
The Janus Faced Scholar: A Festschrift in Honour of Peter Ingwersen. Royal School of Library and Information Science, Copenhagen, pp.325-338, 2010, http://www.academia.edu/attachments/4135543/download_file,

[14] Prpić, K.: *Beyond the Myths about the Natural and Social Sciences: A Sociological View*.
Institute for Social Research Zagreb, Zagreb, 2009,

[15] Newman, M.; Barabasi, A.-L. and Watts, D.J., eds.: *The Structure and Dynamics of Networks*.
Princeton University Press, Princeton and Oxford, 2006,

[16] El-Kholy, E. and El-Esawy, E: *Graph Folding of Some Special Graphs*.
Journal of Mathematics and Statistics **1**(1), 66-70, 2005, http://dx.doi.org/10.3844/jmssp.2005.66.70,

[17] Alvarez-Hamelin, J.I.; Dall'Asta, L.; Barrat, A. and Vespignani, A.: *Large scale networks fingerprinting and visualization using the k-core decomposition.*
In Weiss, Y.; Schölkopf, B. and Platt, J., eds.: Advances in Neural Information Processing Systems **18**. MIT Press, Cambridge, pp.41-50, 2006. http://cnet.fi.uba.ar/ignacio.alvarez-hamelin/pdf/AH_D_B_V_NIPS2006.pdf.

O ČEMU PIŠU HRVATSKI ZNANSTVENICI? ANALIZA DRUŠTVENE I KONCEPTUALNE MREŽE HRVATSKE ZNANSTVENE BIBLIOGRAFIJE

M. Schatten

Fakultet organizacije i informatike – Sveučilište u Zagrebu
Varaždin, Hrvatska

SAŽETAK

U radu se analizira Hrvatska znanstvena bibliografija (CROSBI) kao društvena aplikacija putem nekoliko različitih pristupa. Analizom i vizualizacijom konceptualne mreže utvrđuju se jezgra ključnih riječi za svako od znanstvenih područja: biomedicinske, biotehničke, društvene, humanističke, prirodne i tehničke znanosti. Interpretacijom koncepata iz tako definiranih jezgri putem metapodataka iz društvene aplikacije *Wikipedia* zaključuje se o nesrazmjeru između dvaju društvenih sustava: hrvatske znanstvene zajednice i javnosti. Analizom društvene mreže znanstvenih područja prema društvenom (znanstvena suradnja) i konceptualnom (povezane ključne riječi) kriterijima zaključuje se o neusklađenosti interdisciplinarnih istraživanja.

KLJUČNE RIJEČI

hrvatska znanstvena bibliografija, analiza društvene mreže, analiza konceptualne mreže, vizualizacija, *Wikipedia*, teorija sustava