Laurea University of Applied Sciences
Ratatie 22, 01300 Vantaa, Finland
Phone +358 (0)9 8868 7150
Fax +358 (0)9 8868 7200
firstname.surname@laurea.fi
www.laurea.fi
Business ID 1046216-1
Domicile Vantaa

An Evolutionary Network Analysis of Crowdsourcing Research Community

Teemu Santonen

Laurea University of Applied Sciences
Vanha maantie 9, 02650 Espoo, Finland
E-mail: teemu.santonen@laurea.fi

Mokter Hossain

Aalto University
School of Science
Otaniementie 17, 02150 Espoo, Finland
E-mail: mokter.hossain@aalto.fi

Henri Simula

Aalto University, School of Science
Otaniementie 17, Espoo, Finland
E-mail: henri.simula@aalto.fi

Abstract: Crowdsourcing is an approach to delegate a task to a large group of people, which then suggests their own solutions for the defined task. The objective of this study is to model and describe the structure of the crowdsourcing research community and topics from the social network analysis (SNA) point of view. By applying systematic review we found 493 crowdsourcing publications which were made by 1102 authors. We observed degreasing annual growth in terms of number of publications and authors, resulting leveling interest towards crowdsourcing among scholars. Collaboration within the diffuse crowdsourcing co-author network was tighten only slightly. The structure of author and keyword network differed significantly in term of number of components. Among 1050 co-authors we found 237 subgroups, while in the case of 1243 keywords we identified only 61 interlinked research themes including the substantially large main component which covered 84 per cent share of all keywords.

Keywords: Crowdsourcing, Social Network Analysis, Co-authorship, Keyword

1 Introduction

Crowdsourcing has gained an interest among academics and practitioners ever since the term was coined by Howe (2006). A definition for crowdsourcing has been recently provided by Estellés-Arolas and González-Ladrón-de-Guevara (2012) who found and analysed 40 original and partially conflicting definitions. In addition Zhao and Zhu (2012) conducted an extensive critical examination of existing crowdsourcing research

from information system perspective (IS) and introduced a conceptual framework to prescribe main research objectives of crowdsourcing. In crowdsourcing a task is delegated to a large group of people, which then suggests their own solutions for the defined task (Howe, 2006). Some authors believe that crowdsourcing is just a buzzword and are very skeptic towards it, while others see high potential of it to accomplish various activities (e.g. Poetz and Schreier, 2012, Whitla, 2009, Parameswaran and Whinston, 2007, Eagle, 2009).

However, it seems that the theoretical knowledge base of crowdsourcing is still in its infancy. As a result several uncharted research directions have been suggested such as social issues of crowdsourcing (Brabham 2008, Whitla 2009). Thus, there is a research gap, which this study aims to fill by examining crowdsourcing phenomenon from the social network analysis (SNA) viewpoint. In other words, the purpose of this study is to get insight on crowdsourcing publications and relationship amongst authors via SNA.

This paper is organised as follows. Firstly, we introduce theoretical foundations of crowdsourcing. Secondly, we present our research design including research objectives, data collection, social network analysis as a method and introduce the key variables. Thirdly, we present our results based on network analysis. Finally, we conclude with our findings.

2 Theoretical foundations of crowdsourcing

2.1 Crowdsourcing as a part of online social network and open innovation phenomenon

When wide range of people and their different but complimentary insights are brought together, novel ideas generated by thinking outside the box are possible (Santonen, 2009). Online Social Networks (OSNs) (also known as a social media) including crowdsourcing services are generally referring to communities and hosted online services enabling collaboration (Cachia et. al. 2007) and the creation and exchange of user-generated content in which the consumer is the creator, consumer and distributor of publically available content (e.g. OECD 2007, Le Borgne-Bachschmidt et. al. 2009). In the context of innovation management, user-generated content can be related to user-driven innovation strategy, which emphasizes user's involvement in all development phases from initial idea generation to final launch and production. All this is a part of boarder open innovation strategy approach, which combines internal and external ideas as well as internal and external paths to market to seek business success (Chesbrough, 2003). Basically all these strategies suggest that instead of doing everything by yourself, you should look also for help from external resources e.g. via crowdsouricng. This is important since the network economy believers associate business success with the ability to co-operate with external resources and the circulation of know-how (Pöyhönen and Smedlund, 2004). In context of information systems (IS) similar kind of collaboration approach is typically referring to open source software (www.opensource.org). Open source does not only mean access to the source code, but also e.g. users' freedom to run, copy, distribute, study, change and improve the software depending on the terms of use. A large group of people working independently to achieve shared outcomes through communication technologies and loose voluntary networks is also known as mass collaboration (Tapscott and Williams, 2006) or mass innovation

(Santonen, 2012). It is evident that without OSNs and supporting information systems this kind of crowd or mass collaboration would not be possible.

2.2 Introducing the body of knowledge on Crowdsourcing

The literature review based crowdsourcing studies have clear limitations from innovation body of knowledge viewpoint. The conceptual framework proposed by Zhao and Zhu (2012) approaches crowdsourcing mainly from IS perspective, whereas Estellés-Arolas and González-Ladrón-de-Guevara (2012) focused only to the definition and contextual aspects. Other crowdsourcing studies have been explored various individual issues and that are discussed next.

Usefulness of crowdsourcing for idea generation and problem solving has been discussed for instance by Whitla, (2009), and Poetz and Schreier (2012). Crowdsourcing for problem solving in public and private sectors is also studied (c.f. Brabham 2008, and Jeppesen and Lakhani 2010). There are also studies regarding crowdsourcing in content distribution, marketing and advertising purposes (e.g. Parameswaran and Whinston, 2007 and Whitla, 2009). In addition, the role of users in innovation process is discussed for example by Fuchs and Schreier (2011).

There are several advantages crowdsourcing can bring along but there are also disadvantages (Simula, 2013). One issue with crowdsourcing is the question of how to attract a large number of individuals who has time, willingness, and skills to contribute towards crowdsourcing initiatives. The incentives to do so have been discussed for instance by Eagle (2009), Feller et al., (2012), Zichermann and Linder (2010).

Motivations and reasons to participate in crowdsourcing is discussed by Antikainen et. al., 2010; Brabham, 2010; Hossain, 2010). The motivations can be intrinsic or eccentric but typically monetary rewards are used to motivate participant. According to Feller et al., (2012) presence of rewards increases future participation rate, and absence of it may lead to have negative effects. Terwiesch and Xu (2008) believe that rewarding the best solution with a fixed amount is optimal when quality of the solution depends on many unknown factors.

Apart from typical incentive, a growingly used incentive is utilization of gamification (Zichermann and Linder, 2010). The concept aims to create a task entertaining similar to online games and it helps to engage crowd to perform tasks through gaming. One specific domain of where crowdsourcing has been utilized is distribution of simple tasks to many individual what is labelled as microtasking (Howe, 2008; Eagle, 2009). Additional application of crowdsourcing is to use it to raise funding for various projects i.e. crowdfunding (Lambert and Schwienbacher, 2010; Belleflamme et al., 2010).

Additional research avenue is to investigate crowdsourcing in business-to-business domain (e.g.. Simula et. al. 2012; Simula and Vuori, 2012). There are also discussion on success and failure of crowdsourcing initiatives (Schonfeld, 2008; Simula, 2013),

2.3 Amazon Mechanical Turk as an example of Crowdsourcing service

In order to illustrate possibilities of crowdsourcing Amazon Mechanical Turk is presented as an classical example of crowdsourcing and microtasking. Amazon Mechanical Turk (MTurk**)** was launched in November 2005 (Ross et al. 2010). According to Paolacci et al. (2010) MTurk is "a crowdsourcing web service that co-ordinates the supply and the demand of tasks that require human intelligence to

complete." (p. 411). In practice this means that human beings still can do many tasks more effectively than computers (Iperoitis 2010). According to Amazon these tasks are for instance identifying objects in a photo or video, performing data de-duplication, transcribing audio recordings, or researching data details. (Amazon, 2012). The MTurk workers come mainly from developing countries - for the demographic distribution of worker see for example Ross et al. (2010). This has created some ethical issue Fort et al. (2011) Additional issue with MTurk is that how to verify the quality of the submitted results (Ipeirotis et al. 2010).

3 Research Methodology

3.1 Research Design

The goal of this study is to model and describe the structure of the crowdsourcing research domain from the social network analysis (SNA) point of view (e.g. Wasserman and Faust, 1994). *First*, by applying systematic bibliographic analysis and standard methodology of SNA (Borgatti et al., 1992), we identify who are the key authors based on co-authorship relations and what kind of ties have been constructed among these participants. *Second* by extending SNA also to keywords (e.g. Motter et. al. 1999, Hori et. al, 2004), we are explicitly modelling the different crowdsourcing themes and their inter-linkage with each other. *Third*, by examining author and keyword networks from temporal evolution point of view, we are identifying the historical evolution of crowdsourcing research community.

3.2 Data Collection and Sample Selection

The unit of analysis in this study is a scientific crowdsourcing publication. Following systematic bibliographical search strategy was conducted in order to identify all the relevant studies within our research focus. *Data sources*. The crowdsourcing is a broad and infancy concept. Thus, besides innovation management scholars, it has gain an interest among multiple other disciplines, especially in the field of information system (IS). Therefore in data source selection, we followed similar strategy as Zhao and Zhu (2012) and did not limited the databases by the academic disciplines. As a result following scientific journal databases were added to our bibliographical search: 1) IEEE, 2) ScienceDirect, 3) Emerald, 4) ACM, 5) SAGE, 6) Springer, 7) Wiley, 8) InderScience, 9) Taylor & Francis and 10) World Scientific.

Search terms and fields. In order to identify truly relevant crowdsourcing publications, in the final dataset we included only the publications, which had *crowdsourcing* (or some relevant inflection) as a keyword. Only some of the above introduced databases allowed keyword–field specific search. Therefore, if the word "crowdsourcing" was found in the title, abstract or keyword list of the publication, it was included in the preliminary dataset. Furthermore only the publications in English language approved.

The final selection of publications. After cleaning up the data and correcting obvious spelling errors and unifying plural and unit wordings within all the keywords and author names, we end-up having 497 publications, which originally had *crowdsourcing* term as a keyword. On top of these publications, 25 publications included somehow differently

spelled *crowdsourcing* term as a keyword. Typically these variations were based on combining crowdsourcing with some other term(s) either before or after the crowdsourcing term. The combined variations included such terms as process, innovation, service, data, mobile, enterprise, implicit, platform, evaluation, quality and social web. In few cases we observed *crowd sourcing* spelling instead of *crowdsourcing*. In order to unify the dataset for the further network analysis, alternative keyword spellings were replaced by the simplified *crowdsourcing* term. Next, duplicate cases were identified with the help of SPSS by comparing the titles of the publications and then verifying that also authors and keywords were the same. In one of the duplicated cases, the number of authors had increased by one. In few cases the order of the keywords had changed a little bit. Since the keyword order does not effect on the relationship result, this difference was bypassed. In three publications, the number of keywords had increased by one keyword. In those cases the publication having greater amount of keywords was selected. As result 29 publications appeared to be duplicated case (i.e. same title, authors and keywords). Thus, them were included only once in the dataset. In all we had now 493 valid and individual publications for our analysis.

Since only some of the selected databases allowed keyword–field specific search functionality, additional fields were used during the search process. Therefore original search results included also 21 publications which had keywords, but within those keywords crowdsourcing (or some kind of variation) term was not detected. Moreover, we also found 63 publications which did not include keywords at all. In the case of 84 publications without crowdsourcing keyword, 52 of them (i.e. 62 per cent) include crowdsourcing term in headline. However, in order to ensure comprehensive and solid dataset including both author names and keywords, we omitted these publications from the further analysis. This filtering decision must be considered as a potential error source.

3.3 Social Network Analysis as a method

An extensive body of knowledge regarding social network studies exist (e.g. Wasserman and Faust 1994, Scott 2000, Watts 1999) while the advantages of researchers' collaboration has also been popular topic among scholars (e.g., Beaver and Rosen 1978, 1979a, 1979b, Melin, 2000, Autry and Griffis 2005, Kuhn 1962, Barnett et. al. 1988). Due to the benefits of the collaboration, there is a tendency among scholars to collaborate with other researchers or practitioners. SNA studies can help in this partner finding process by offering empirical and visual data on the linkage between researchers and topics including an evolution pattern of a particular field of study. On the contrary to non-network-type of studies, social network analysis study includes relevant relational information beyond the attributes of individual actor which helps understanding individuals and communities behaviour relating to their social structures (Yan and Assimakopoulos, 2009). Previously SNA has been successfully used to study various types of scientific communities (e.g. Newman 2001, Morlacchi et. al. 2005, Vidgen et. al. 2007) including global open innovation research (Su and Lee 2012) and a network structure of ISPIM community (Santonen and Ritala 2012). These studies have demonstrated the usefulness of SNA to reveal underlying structures of scientific domain. Until now, SNA method has not yet been applied to the crowdsourcing research community.

3.4 Key Measures

In SNA studies a *node* can refer to any kind of actor within a network. In this study it refers to an author or a keyword. In turn, centrality measures help in determining the importance of a particular node in the network (Wassermann and Faust, 1994). To indicate centrality of nodes, we utilise measures of *degree centrality* and *betweenness centrality*.

Degree centrality (Freeman, 1979) is the most common measure to evaluate and interpret the node's network position. Degree centrality calculates how many direct connections each node has with other nodes in the network. Thus, it show how linked each node is to other nodes. A high degree centrality then indicates that the node has a central position in the network among other nodes (indicating e.g. a "hub" or otherwise relevant position).

Betweenness centrality (Freeman, 1979) is used for investigating the structural position of a particular node between clusters of nodes in a network. Therefore it can be interpreted as measuring the nodes based on their position and role as a gatekeeper between two or more independent components. Such nodes may be in a structurally powerful position because they might be able to exploit their gatekeeper role for the purposes of knowledge and resource sharing between the separate parts of the network, for example.

We will also conduct an analysis of *network components*. According to Hawe et. al. (2004), a component is a part of a network in which all authors are directly or indirectly connected by at least one connection. Thus, the component analysis will reveal those groups within the whole of the crowdsourcing research community that are internally connected, but separate from each other.

4 Results

4.1 Annual distribution of crowdsourcing publications

In Figures 1 and 2 we have presented the annual distribution and growth of our 493 crowdsourcing publications. The more detailed descriptive statistics are presented in Appendix 1, Table 1.

Figure 1. Number of crowdsourcing publication from 2007 to 2012

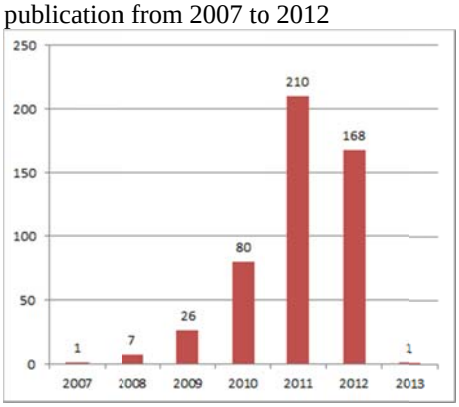
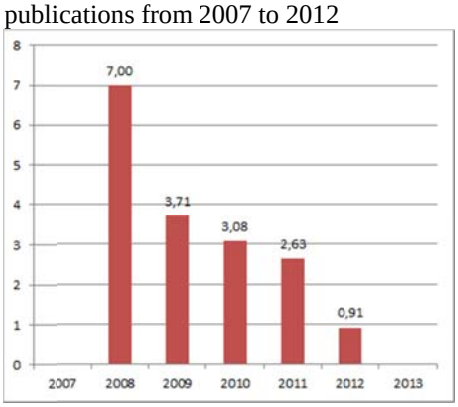


Figure 2. Growth rate of crowdsourcing publications from 2007 to 2012

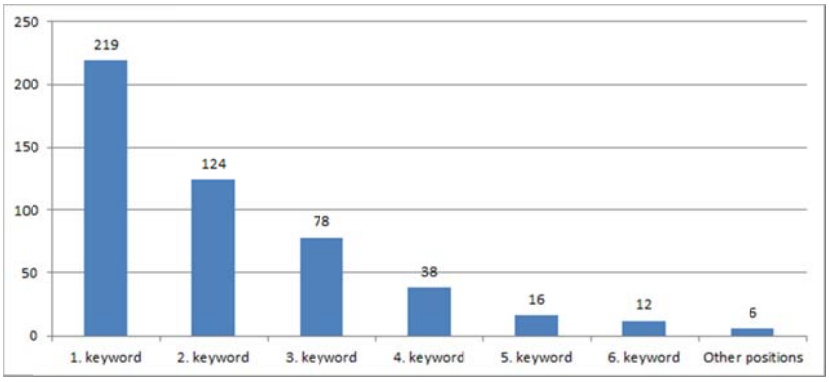


According to Table 1 only limited number (N=26) of crowdsourcing papers were published between 2007 to 2009 resulting 6.9 cumulative per cent share when compared to our full dataset of 493 publications. From 2009 to 2010 we observed significant increase from 26 to 80 annual publications. Likewise, the amount of annual publications increased significantly from 80 to 210 during 2010 to 2011. Interestingly, we detected a drop in the number of publications from 2011 to 2012. Since our data collection was finished in October 2012, this might partially explain this finding. Obviously, some of the publications for 2012 were not yet available due to unfinished year. On the other hand we might be observing a decreasing or at least a leveling interest towards crowdsourcing among scholars. If we assume that the studies are published evenly throughout the year, based on our figures, the annual growth from 2011 to 2012 will remain close to one or can even be negative (i.e. growth rate less than 1). As an interesting specialty, we must point out Crowdsourcing and Ethics publication by Harris and .Srinivasan (2013), which was dated to 2013, even if our data was collected in 2012.

4.2 Relative location/importance of crowdsourcing keyword

In order to verify the relevance of our dataset, we conducted a keyword analysis which evaluated the relative location of the crowdsourcing keyword with in the publications keyword list (Appendix 1, Table 2 and Figure 3). It was assumed that the higher the crowdsourcing keyword position is, the better the publication is focusing on the crowdsourcing topic.

Figure 3. Relative location of the crowdsourcing keyword (N=493)

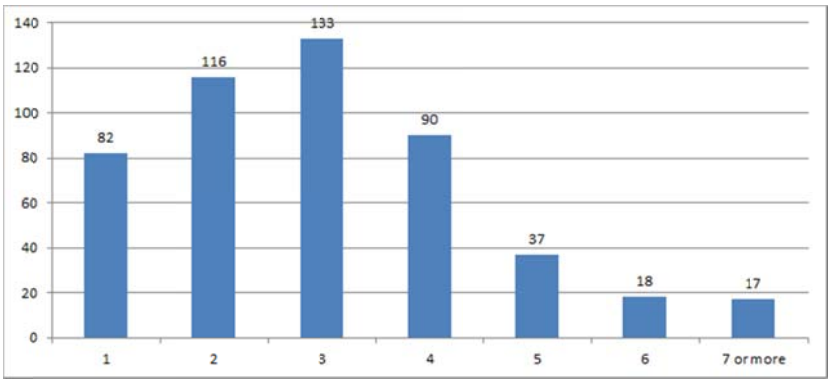


The maximum amount of keywords per one paper was as high as sixteen. In the case of 219 publications (i.e. 44.4 per cent of all publications), the crowdsourcing keyword was the first keyword. In the case of 124 publications, crowdsourcing was the second keyword resulting 69.6 cumulative per cent when calculated together with the first keyword. Cumulative per cent exceeded 85 per cent, when three keywords were included. As a result we argue that the significant share of the publications in our dataset is genuinely focusing on the crowdsourcing topic and therefore can be used for further analysis.

4.3 A Descriptive Profile of crowdsourcing author network

In Appendix 1, Table 3 we have presented a distribution of number of authors per publications, while in Figure 4 (see next page) we have visualised this distribution. Majority of the publication were co-authored (83.4 per cent share). Most typically publications included two (23.5 per cent), three (27 per cent) or four (18.3 per cent) co-authors resulting 68.8 per cent cumulative share. Publications with 5 or more authors were nearly as common as individually written publications (14.6 per cent vs. 16.6 per cent). Thus, for the co-author network analysis we will have all together 411 publications.

Figure 4. Number of authors per publication



In all our data set included 1050 different authors, which had co-authored at least one publication. In addition our data set included 82 single authored publications which were made by 71 different authors. From these 19 authors (27 per cent) had also co-authored some other publication. Thus, those 52 authors who did not co-author any papers were in practice excluded from the further social network analysis.

In order to evaluate the expansion of the crowdsourcing co-author network, a temporal analysis of the network size and connections was conducted. The more detailed descriptive statistics are presented in Appendix 1, Table 4 and visualized in figures 5 and 6 (see next page).

Figure 5. Size of the co-author network between 2007 to 2012

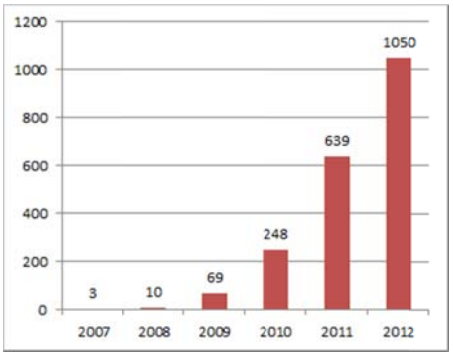
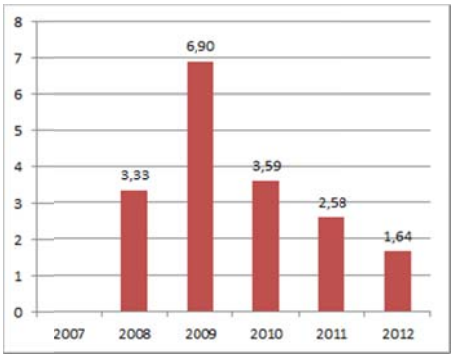


Figure 6. Annual growth rate from 2007 to 2012



As the number of publications results already evidently indicated, the co-author network of crowdsourcing scholars was rather limited from 2007 to 2009. By 2009 the network included 69 authors, which represented only 6.6 per cent share if compared to the size of 2012 crowdsourcing network. So far we have identified clear annual growth regarding the co-author network size, though the growth rate has past three years been dropping from 3.59 to 1.64.

4.4 A Component analysis of crowdsourcing author network

In order to understand the co-author network structure more deeply, we conducted a component analysis which reveals the separate co-authorship groups within the crowdsourcing scholar community. In Figure 7 we have presented crowdsourcing co-author community component structure in year 2012, in which the main component is visualised with the dark grey. The total number of components within crowdsourcing community is 237. In Figure 8 we have visualized the distribution of the number of authors per component in year 2012, while in Appendix 2 Table 1 we have presented the detailed descriptive statistics.

Figure 7. Crowdsourcing co-author network component structure where the main component is visualised with the darker grey.

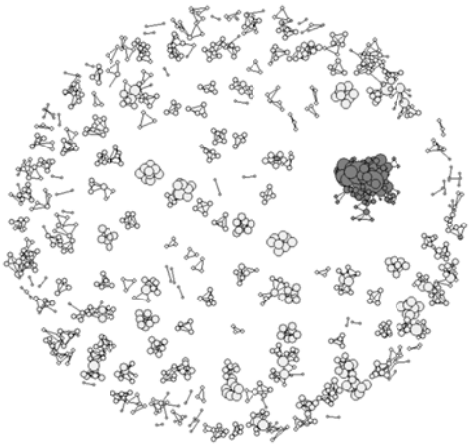
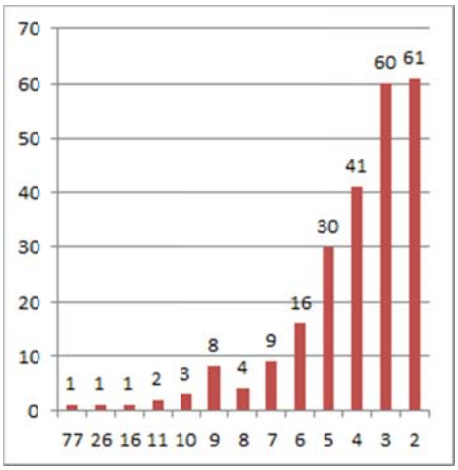


Figure 8. The distribution of the number of authors per component in year 2012



Altogether the crowdsourcing community included 1050 different authors who had co-authored at least one publication. As Figure 7 and 8 illustrates, the majority of components are very small (i.e. over 50 per cent includes only two or three authors). The main component included 77 authors representing 7.3 per cent share of the total co-author population, while second largest component included significantly less, only 26 (= 2.5 per cent share) and third largest only 16 (=1.5 per cent share) authors. Thus, our result indicate rather diffuse network of scholar, who are interested in crowdsourcing.

To gain more in-depth understating regarding the evolution of crowdsourcing co-author network structure, in Figure 9 we have visualised the number of authors (N) and components (C) and connections from 2007 to 2012. Moreover, in Figures 10 we have compared the annual number of authors and components, while in Figure 11 we have

presented authors per component ratio. As a result, we identified only slightly tighten collaboration among crowdsourcing scholars (i.e. the number of authors per component is slowly increasing).

Figure 9. Expansion of the crowdsourcing co-author network

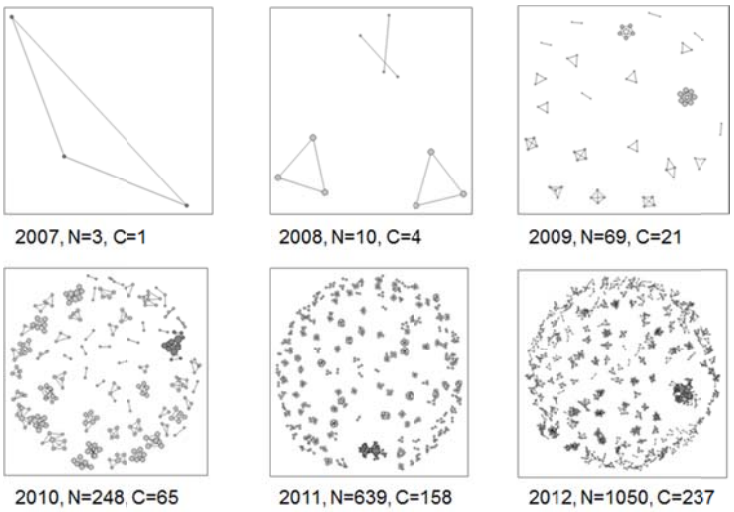


Figure 10. Comparison of number of authors and component from 2007 to 2012

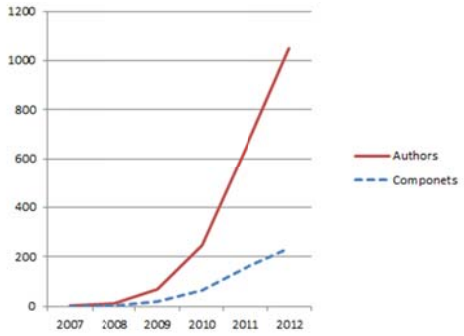
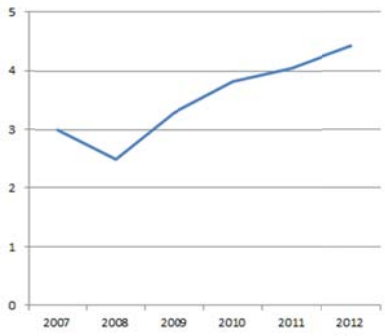


Figure 11. Authors per component ratio from 2007 to 2012



4.5 An ego-analysis of co-author network

When the detailed ego analysis of crowdsourcing co-author network was conducted (Appendix 2, Table 2) following observations were made. Scholars from USA seem to dominate the top places when betweenness centrality is used as a ranking measure since 19 out 25 authors come from USA. Other countries include UK (2 authors), Germany (2), Austria (1) and Israel (1). Members of the main component seem to do well due taking the first eight places on the top ranking list. Based on the betweenness centrality measure, the dominant authors are Hartmann (1153), Miller (1020) and Ackerman (892). In the case of the degree centrality, we had similar results as in the case of betweenness centrality. Authors from USA are also dominating the top ranks (7 authors out of top 10

are from USA). The remaining 3 authors come from Austria. Interestingly, the Information System (IS) scholars seemed to overrule innovation management scholars if the top rankings with in betweenness and degree centrality are used as metrics. In our data set, about 63 per cent of the data was coming from the Association for Computing Machinery (ACM) publications, which is the world's largest educational and scientific computing society.

4.6 A Component and ego-analysis of relating keyword network

IS scholars domination within the co-author network raised our attention regarding the crowdsourcing research topics. Like in the case of co-author network, we conducted a component analysis to keywords in order to reveal the hidden sub groups (Appendix 3, Table 1). In Figure 13 we have visualised resulting component structure which includes 61 components. As the figure clearly show, the relating keyword network structure differs a lot comparing co-author network. We identified a large main component which had 1039 keywords resulting 84 per cent share of all keywords. The remaining keyword 60 components included only 2 to 9 keywords. In Figure 14 we have visualised all the main component nodes which degree was 35 or greater due to otherwise too blurry and small to read network image. This helped us to amplify the balance and inter linkage between most important keyword nodes and search explanation of IS scholars domination.

Figure 13. Crowdsourcing keyword network component structure where the main component is visualised with the red

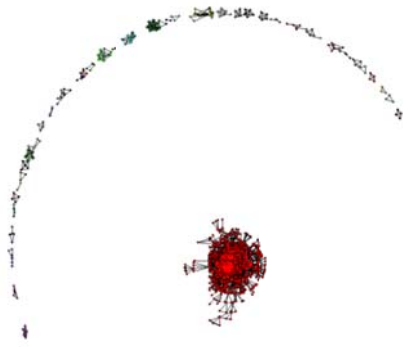
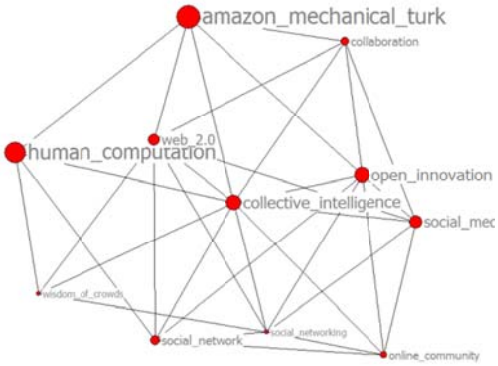


Figure 14. The main keyword component including nodes having degree 35 or more



The most diversely connected keyword in terms of betweenness and degree centrality was Amazon Mechanical Turk keyword with 172 connections. During the keywords filtering process we actually identified multiple different wordings for *"Amazon Mechanical Turk"*, which however during the data clean-up process were unified. In order to explain why this particular keyword gained the top position, the in-depth content analysis of the publications having Amazon Mechanical Turk as a keyword, was conducted. As a result of our analysis it seems that Amazon Mechanical Turk is a good tool for instance to run surveys and thereby ensures a great response (i.e. people get paid if they will participate but it is still possible to do without large budget as payments are typically very low). Thus, the crowdsourcing in that sense is just a method to conduct

research, not a research subject as such. To conclude this finding we suggest that crowdsourcing approach provides innovative way to conduct research, even if those publications do not study crowdsourcing as phenomenon per se. Our observation is in-line with previous suggestions by Burmester at. al 2011 who argued that crowdsourcing is a fast tool to obtain high-quality data inexpensively or a tool to conduct all kinds of experiments Paolacci et al. (2010). Besides Amazon Mechanical Turk application, Twitter, Wikipedia and undefined "location based services" were the best ranking keywords when only application names are considered.

The second ranking keyword with 161 connections was human computation, which instead of innovation management is more related to Information System (IS) domain. Furthermore, besides the human computation keyword we identified other more IS domain related terms such as open source, online community, mobile/mobile phone, data mining, sematic web, machine learning and video. This finding gave more support to our previous suggestion that IS scholar and topics appears to dominate within the crowdsourcing community. However, in further studies this suggestion should be verified statistically by doing additional classification analysis between IS and Innovation management scholars. As a result it has been natural progress that the major conceptual framework efforts regarding crowdsourcing have become from IS field (Zhao and Zhu, 2012) instead of innovation management.

Third, the keyword related ego analysis revealed that from theoretical point of view scholar clearly see the crowdsourcing as a part of larger social media, web 2.0, social networking/network phenomenon. We want to remark that in our keyword analysis we did not unite e.g. social network or social networking term, which would have even more amplify these keywords presence among crowdsourcing research themes.

5 Conclusions

In order to describe the structure and development path of the crowdsourcing research community and themes, a bibliographic analysis grounded social network analysis (SNA) was conducted. With the help of standard methodology of SNA and various additional descriptive statistical analysis we identified the co-author network structure and main research themes and the inter-linkage of these nodes from the temporal evolution point of view. We also described and compared the author and keyword network structures with each other. As a result we were able to provide comprehensive outlook to the crowdsourcing community in its current and historical state.

Our results are limited since this study covered only author and keyword viewpoints. Moreover, the bibliographical search and the followed publication filtering process included in the dataset only the publication which had crowdsourcing term as a keyword excluding all other possible crowdsourcing studies from the analysis. In addition, we were using the crowdsourcing keyword and it's position among publication's keyword list as relevance validator of publication. To ensure the genuine relevance of the publication as crowdsourcing publication, one should do more detailed content analysis and include/exclude publication on the basis of manual selection. Moreover, the future studies should evaluate also the keyword component structure from evolutionary point of view and compare these results with the co-author network structure evolution.

References

Amazon, (2012), Retrieved November 15, 2012, from https://www.mturk.com/mturk/help?helpPage=overview

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk A New Source of Inexpensive, Yet High-Quality, Data?. Perspectives on Psychological Science, 6(1), 3-5.

Fort, K., Adda, G., & Cohen, K. B. (2011). Amazon mechanical turk: Gold mine or coal mine?. Computational Linguistics, 37(2), 413-420.

Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on amazon mechanical turk. Judgment and Decision Making, 5(5), 411-419.

Ross, J., Irani, L., Silberman, M., Zaldivar, A., & Tomlinson, B. (2010, April). Who are the crowdworkers?: shifting demographics in mechanical turk. In Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems (pp. 2863-2872). ACM.

Ipeirotis, P. G. (2010). Analyzing the amazon mechanical turk marketplace. XRDS: Crossroads, The ACM Magazine for Students, 17(2), 16-21.

Ipeirotis, P. G., Provost, F., & Wang, J. (2010, July). Quality management on amazon mechanical turk. In Proceedings of the ACM SIGKDD workshop on human computation (pp. 64-67). ACM. Amazon, (2012), Retrieved November 15, 2012, from https://www.mturk.com/mturk/help?helpPage=overview

Antikainen, M. J., & Vaataja, H. K. (2010). Rewarding in open innovation communities–how to motivate members. International Journal of Entrepreneurship and Innovation Management, 11(4), 440-456.

Autry, Chad W., and Stanley E. Griffis (2005), "A Social Anthropology of Logistics Research: Exploring Productivity and Collaboration in an Emerging Science." Transportation Journal, 44 (4), pp. 27-43.

Barnett, A., Ault, R.W., Kaserman, D.L., (1988), A Social Anthropology of Logistics Research: Exploring Productivity and Collaboration in an Emerging Science, Transport Journal, Vol. 44, No. 4, pp. 27-44

Beaver D, Rosen R. (1978), Studies in scientific collaboration: Part I. The professional origins of scientific co-authorship. Scientometrics, Vol. 1, pp. 65–84.

Beaver D, Rosen R. (1979a), Studies in scientific collaboration: Part II. Scientific co-authorship, research productivity, and visibility in the French scientific elite. Scientometrics, Vol. 1, pp. 139–149.

Beaver D, Rosen R. (1979b), Studies in scientific collaboration: Part III. Professionalization and the natural history of modern scientific co-authorship. Scientometrics, Vol. 1, pp. 231–245.

Belleflamme, P., Lambert, T., & Schwienbacher, A. (2010). Crowdfunding: tapping the right crowd. Discussion Paper, Center for Operations Research and Econometrics.

Borgatti, S.P. and Everett, M.G. (1999) Models of core/periphery structures. Social Networks, Vol. 21, pp. 375-395.

Borgatti, S.P., Everett, M.G. and Freeman, L.C. (1992). Ucinet – Guide – Ucinet for Windows: Software and Social Network Analysis. Harvard, MA: Analytic Technologies.

Brabham, D. C. (2008). Moving the crowd at iStockphoto: The composition of the crowd and motivations for participation in a crowdsourcing application. First Monday, 13(6), 1-22.

Brabham, D. C. (2010). Moving the crowd at Threadless. Information, Communication & Society, 13(8), 1122-1145.

Cachia, R., Compañó, R., Da Costa, O., (2007), Grasping the potential of online social networks for foresight, Technological Forecasting and Social Change 74(8), pp. 1179-1203.

Chesbrough, H. W. (2003). Open innovation: The new imperative for creating and profiting from technology. Harvard Business Press.

Eagle, N. (2009). txteagle: Mobile crowdsourcing. Internationalization, Design and Global Development, pp. 447-456.

Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. Journal of Information science, 38(2), 189-200.

Feller, J., Finnegan, P., Hayes, J., & O'Reilly, P. (2012). 'Orchestrating'sustainable crowdsourcing: A characterisation of solver brokerages. The Journal of Strategic Information Systems.

Fort, K., Adda, G., & Cohen, K. B. (2011). Amazon mechanical turk: Gold mine or coal mine?. Computational Linguistics, 37(2), 413-420.

Freeman, L.C. (1979). Centrality in networks: Conceptual clarification, Social Networks, Vol. 1, pp.215–239.

Fuchs, C., & Schreier, M. (2011). Customer empowerment in new product development. Journal of Product Innovation Management, 28(1), 17-32.

Garrick 2009,

Hawe, P, Webster, C., and Shiell, A. (2004). A glossary of terms for navigating the field of social network analysis. J Epidemiol Community Health, Vol. 58, pp. 971–975.

Hori, K., Nakakoji, K., Yamamoto, Y., & Ostwald, J. (2004). Organic perspectives of knowledge management: Knowledge evolution through a cycle of

knowledge liquidization and crystallization. Journal of Universal Computer Science, 10(3), 252–261.

Hossain, M. (2012). Users' motivation to participate in online crowdsourcing platforms. In Innovation Management and Technology Research (ICIMTR), 2012 International Conference, pp. 310-315. IEEE.

Howe, J. (2006). The rise of crowdsourcing. Wired magazine, 14(6), 1-4.

Ipeirotis, P. G. (2010). Analyzing the amazon mechanical turk marketplace. XRDS: Crossroads, The ACM Magazine for Students, 17(2), 16-21.

Ipeirotis, P. G., Provost, F., & Wang, J. (2010, July). Quality management on amazon mechanical turk. In Proceedings of the ACM SIGKDD workshop on human computation (pp. 64-67). ACM.

Jeppesen, L. B., & Lakhani, K. R. (2010). Marginality and problem-solving effectiveness in broadcast search. Organization Science, 21(5), 1016-1033.

Kuhn, T.S., (1962). The Structure of Scientific Revolutions, 1st. ed., Chicago: Univ. of Chicago Pr.

Lambert, T., & Schwienbacher, A. (2010). An empirical analysis of crowdfunding. Louvain–la–Neuve: Louvain School of Management, Catholic University of Louvain

Le Borgne-Bachschmidt, F., Girieud, S., Leiba, M., de Munck S., Limonard, S., Poel, M., Kool, L., Helberger, N., Guibault, L., Janssen, E., Van Eijk, N., Angelopoulos, C., Van Hoboken, J., Swart, E. (2009), User Generated Content: Supporting a participative Information, Society Study for the European Commission DG Information Society, Final Report

Leimeister, J. M. (2010). Collective intelligence. Business & Information Systems Engineering, 2(4), 245-248.

Melin, G., (2000). Pragmatism and Self-organization Research Collaboration on the Individual Level, Research Policy, Vol. 29, No.1, pp.31-40.

Morlacchi, P., Wilkinson, I. F. and Young, L. C. (2005). Social networks of researchers in B2B Marketing: A case study of the IMP Group 1984–1999. Journal of Business-to-Business Marketing, Vol. 12, pp. 3–34.

Motter, A. E., de Moura, A. P. S., Lai, Y. C., & Dasgupta, P. (1999). Topology of the conceptual network of language. Physical Review E, 65, 065102.

Newman, M. E. (2001). The structure of scientific collaboration networks. Proc NatlAcad Sci U S A, Vol. 98, pp. 404–409.

OECD, (2007), Participative Web: User-created Content, DSTI/ICCP/IE(2006)7/FINAL, prepared by Sacha Wunsch-Vincent and Graham Vickery, Directorate for science, technology and industry; Committee for information, computer and communications policy

Paolacci, G., Chandler, J., & Ipeirotis, P. (2010). Running experiments on amazon mechanical turk. Judgment and Decision Making, 5(5), 411-419.

Parameswaran, M., & Whinston, A. B. (2007). Social computing: An overview. Communications of the Association for Information Systems, 19(37), 762-780

Poetz, M. K., & Schreier, M. (2012). The value of crowdsourcing: can users really compete with professionals in generating new product ideas?. Journal of Product Innovation Management, 29(2), 245-256.

Pöyhönen, A., Smedlund, A., (2004), Assessing intellectual capital creation in regional clusters. Journal of Intellectual Capital, Vol.5, No. 3; pg. 351.

Ross, J., Irani, L., Silberman, M., Zaldivar, A., & Tomlinson, B. (2010, April). Who are the crowdworkers?: shifting demographics in mechanical turk. In Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems (pp. 2863-2872). ACM.

Santonen, T., (2009), Creating the Foundations for Mass Innovation Implementing National Open Innovation System (NOIS) as a Part of Higher Education, In K.R.E. Huizingh, M. Torkkeli, S. Conn and I. Bitran (eds). Proc. Of The 2nd ISPIM Innovation Symposium - Stimulating Recovery - The Role of Innovation Management, New York City, USA on 6-9 December 2009

Santonen, T., (2012), Massidea.org – a greener way to innovate in "Green Technologies in Food Production and Processing", Yves Arcand and Joyce Boye (eds), pp. 541-568, Springer Science+Business Media, LLC, USA

Santonen, T., Ritala, P. (2012), Social Network Analysis of ISPIM Innovation Management Community, in the XXII ISPIM Conference – The XXIII ISPIM Conference – Action for Innovation: Innovating from Experience - Barcelona, Spain

Schonfeld, E. 2008, When Crowdsourcing Fails: Cambrian House Headed to the Deadpool, Available at http://techcrunch.com/2008/05/12/when-crowdsourcing-fails-cambrian-house-headed-to-the-deadpool/ [Accessed on 20,08,2012]

Scott, J. P (2000). Social Network Analysis: A Handbook. SAGE Publications.

Simula, H (2013) The Rise and Fall of Crowdsourcing? Forthcoming. Accepted to be presented at Hawaii international conference on system science.

Simula, H., Töllinen, A., & Karjaluoto, H. (2012). Facilitating innovations and value co-creation in industrial B2B firms by combining digital marketing, social media and crowdsourcing. Proceedings of the 2012 Academy of Marketing Science Annual Conference, 16-19 May, New Orleans, USA.

Simula, H., Vuori, M. (2012). Benefits and Barriers of Crowdsourcing in B2B Firms: Generating Ideas with Internal and External Crowds. Forthcoming. Accpeted to be published in the December issue of International Journal of Innovation Management.

Su, H-N, Lee, P-C, (2012), Framing the structure of global open innovation research, Journal of Informetrics 6 (2012) 202– 216

Tapscott, D., Williams, A.D., (2006), Wikinomics: How Mass Collaboration Changes Everything, Portfolio Hardcover

Terwiesch, C., & Xu, Y. (2008). Innovation contests, open innovation, and multiagent problem solving. Management Science, 54(9), 1529-1543.

Wasserman, S. and K. Faust (1994). Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press.

Watts, D. J. (1999) Small Worlds, Princeton, NJ: Princeton Univ. Press.

Whitla, P. (2009). Crowdsourcing and its application in marketing activities. Contemporary Management Research, 5(1).

Vidgen, R., Henneberg, S. and Naudé, P. (2007). What sort of community is the European Conference on Information Systems? A social network analysis 1993–2005, European Journal of Information Systems, Vol. 16, pp. 5-19.

Yan, J. and Assimakopoulos, D. (2009). The small-world and scale-free structure of an internet technological community. International Journal of Information Technology and Management, Vol. 8, pp. 33-49.

Zhao, Y., & Zhu, Q. (2012). Evaluation on crowdsourcing research: Current status and future direction. Information Systems Frontiers. Published online: 11 April 2012.

Zichermann, G., & Linder, J. (2010). Game-based marketing: inspire customer loyalty through rewards, challenges, and contests (hardback).

Appendix 1

Table 1. An annual distribution of crowdsourcing publications (N=493)

Year	Frequency	Per cent	Cumulative per cent	Annual growth
2007	1	0.2	0.2	
2008	7	1.4	1.6	7.00
2009	26	5.3	6.9	3.71
2010	80	16.2	23.1	3.08
2011	210	42.6	65.7	2.63
2012	168	34.1	99.8	)* 0.80 → 0.91 estimation based on even annual publication frequency
2013	1	0.2	100	

)* NOTE: Data collection was finished in October 2012.

Table 2. Relative location of crowdsourcing keyword (N=493)

Position of the crowdsourcing keyword	Frequency	Per cent	Cumulative per cent
Crowdsourcing as the 1. keyword	219	44.4	44.4
Crowdsourcing as the 2. keyword	124	25.2	69.6
Crowdsourcing as the 3. keyword	78	15.8	85.4
Crowdsourcing as the 4. keyword	38	7.7	93.1
Crowdsourcing as the 5. keyword	16	3.2	96.3
Crowdsourcing as the 6. keyword	12	2.4	98.8
Other positions	6	1.2	100

Table 3. Distribution of number of authors per publication (N=493)

Number of authors	Frequency	Per cent	Cumulative per cent
1	82	16.6	16.6
2	116	23.5	40.2
3	133	27.0	67.1
4	90	18.3	85.4
5	37	7.5	92.9
6	18	3.7	96.6
7 or more	17	3.4	100.0

Table 4. Crowdsourcing co-author network: size and growth.

Years	Total number of authors	Number of new authors	Cumulative share compared to 2012 population	Annual growth
2007	3	3	0.3 %	
2007 – 2008	10	7	1.0 %	3.33
2007 – 2009	69	59	6.6 %	6.90
2007 – 2010	248	179	23.6 %	3.59
2007 – 2011	639	391	60.9 %	2.58
2007 – 2012	1050	411	100.0 %	1.64

Appendix 2

Table 1. Distribution of the number of authors per component in year 2012

Component size (number of authors)	Frequency	Per cent	Total number of authors with in component(s)	Cumulative per cent of the number of authors
77	1	7.3 %	77	7.3 %
26	1	2.5 %	26	9.8 %
16	1	1.5 %	16	11.3 %
11	2	2.1 %	22	13.4 %
10	3	2.9 %	30	16.3 %
9	8	6.9 %	72	23.1 %
8	4	3.0 %	32	26.2 %
7	9	6.0 %	63	32.2 %
6	16	9.1 %	96	41.3 %
5	30	14.3 %	150	55.6 %
4	41	15.6 %	164	71.2 %
3	60	17.1 %	180	88.4 %
2	61	11.6 %	122	100.0 %

Table 2. Crowdsourcing network in year 2012 – Ego analysis 1 of 1050 authors

	Betweenness centrality			Degree centrality	
1	Hartmann B.(USA) [1]	1153.900	1	Hartmann B. (USA) [1]	37
2	Miller Robert (USA) [1]	1020.300	2	Schall Daniel (AT) [5a]	33
3	Ackerman Mark (USA) [1]	892.000	3	Dustdar Schahram (AT) [5a]	30
4	Bernstein M. (USA) [1]	650.167	4	Miller Robert (USA) [1]	26
5	Horton John (USA) [1]	355.000	5	Bernstein M. (USA) [1]	24
6	Kittur Aniket (USA) [1]	348.433	6	Vukovic Maja (USA) [3]	21
7	Bigham Jeffrey (USA) [1]	253.000	7	Ackerman Mark (USA) [1]	21
8	Shaw Aaron (USA) [1]	221.000	8	Psaier Harald (AT) [5a]	17
9	Lease Matthew (USA) [2]	201.000	9	Kittur Aniket (USA) [1]	17
10	Yilmaz Emine (UK) [2]	184.000	10	Kulkarni Anand (USA) [1]	17
11	Horvitz Eric (USA) [1]	148.000	11	10 authors	16
12	Klemmer Scott (USA) [1]	148.000	12	1 author	15
13	Alonso Omar (GER) [2]	125.000	13	1 author	14
14	Chilton Lydia (USA) [1]	122.500	14	3 authors	13
15	Lasecki Walter (USA) [1]	105.000	15	16 authors	12
16	Greenshpan Ohad (IL) [3]	68.000	16	13 authors	11
17	Kazai Gabriella (UK) [2]	66.000	17	7 authors	10
18	Vukovic Maja (USA) [3]	64.000	18	12authors	9
19	Karger David (USA) [1]	61.500	19	55 authors	8
20	Madden S. (USA) [1]	41.367	20	28 authors	7
21	Hoßfeld Tobias (GER) [4a]	31.000	21	94 authors	6
22	André Paul (USA) [1]	20.167	22	75 authors	5
23	Hong Jason (USA) [5b]	20.000	23	188 authors	4
24	Dustdar Schahram (AT) [5a]	19.000	24	188 authors	3
25	Kulkarni Anand (USA) [1]	17.000	25	216 authors	2
26	Garcia-Molina H. (USA) [5c]	16.000	26	133authors	1

NOTE: Country codes: 1) United States of America = USA, 2) Germany = GER, 3) Austria = AT, 4) Israel = IL. Numbers [1 to 5c] are used as a component identifier.

Appendix 3:

Table 1. Distribution of the number of keywords per component

Component size (number of keywords)	Frequency	Per cent	Total number of keywords with in component(s)	Cumulative per cent of the number of keywords
1039	1	1.6 %	1039	83.6 %
9	1	1.6 %	9	84.3 %
8	1	1.6 %	8	85.0 %
7	1	1.6 %	7	85.5 %
6	3	4.9 %	18	87.0 %
5	8	13.1%	40	90.2 %
4	8	13.1 %	32	92.8 %
3	14	23.0 %	42	96.1 %
2	24	39.3 %	48	100.0 %

Table 2. Crowdsourcing network – Ego analysis 1 of 1050 keywords

	Betweenness centrality			Degree centrality	
1	Amazon mechanical turk	19366.1	1	Amazon mechanical turk	172
2	Human_computation	19361.0	2	Human computation	161
3	Collective intelligence	19110.7	3	Collective intelligence	112
4	Open innovation	17755.5	4	Open innovation	106
5	Web 2.0	14369.2	5	Web 2.0	97
6	Social media	14251.6	6	Social media	89
7	Social network	13729.0	7	Social network	64
8	Collaboration	13369.6	8	Online community	60
9	Online community	13129.9	9	Collaboration	54
10	Motivation	12747.1	10	Social networking	48
11	Social computing	12109.1	11	Wisdom of crowds	43
12	Games with a purpose	11766.4	12	Motivation	36
13	Twitter	11302.0	13	Governance	36
14	Crowds	10543.8	14	Smart city	36
15	Wisdom of crowds	10031.0	15	Social computing	35
16	Community	9721.7	16	Co-creation	32
17	Governance	9704.8	17	Mobile	30
18	Innovation	9432.8	18	Open source	28
19	Co-creation	9301.5	19	Games with a purpose	28
20	Privacy	19366.1	20	Incentive	26
21	Data mining	19361.0	21	Innovation	25
22	Crisis informatics	19110.7	22	Twitter	25
23	Coordination	17755.5	23	Data mining	24
24	Location based service	14369.2	24	e-government	24
25	Open source	14251.6	25	Semantic web	23
26	Social networking	13729.0	26	Evaluation	22
27	Quality assurance	13369.6	27	Volunteer geographic inf.	22
28	Wikipedia	13129.9	28	Machine learning	22
29	Reliability	12747.1	29	Crisis informatics	21
30	Mobile phone	12109.1	30	Video	20