

A Social Network Analysis of the Information Fusion Community

Fredrik Johansson, Christian Mårtenson, Pontus Svenson
Swedish Defence Research Agency (FOI)
SE-164 90 Stockholm, Sweden
Email: {frejoh,cmart,ponsve}@foi.se

Abstract—Social network analysis is an important research area for supporting intelligence analysts since it can be used to identify important actors and subgroups in e.g. criminal networks. In this paper, we apply methods from social network analysis on bibliographic data about the research area of information fusion, in order to demonstrate how this kind of algorithms can be applied to get a better overview of the information fusion research community. The results show which authors have published the most in the research community, which authors that are most “powerful” in terms of various metrics widely used in social network analysis, and what the most and least active geographical areas are.

Keywords: Information fusion, Intelligence analysis, SNA, Social network analysis.

I. INTRODUCTION

Information fusion research is becoming increasingly mature. The FUSION 2011 conference in Chicago will be the fourteenth international conference on information fusion, and the research community is steadily becoming stronger, and attracting new researchers into the field. However, as the community grows it also becomes increasingly harder to overlook, in part due to its focus on applied research. New techniques and methodologies are added to and developed within the information fusion domain. For newcomers to the field, it is not easy to get a quick and good overview of the area, and even for experienced information researchers it is hard to clearly separate what is information fusion from what is not. A good overview is given in [1], but the information fusion field has evolved much since then. Definitions of information fusion such as Dasarathy’s [2]:

“Information fusion encompasses the theory, techniques, and tools conceived and employed for exploiting the synergy in the information acquired from multiple sources (sensors, databases, information gathered by human, etc.) such that the resulting decision or action is in some sense better (qualitatively and quantitatively, in terms of accuracy, robustness and etc.) than would be possible, if these sources were used individually without such synergy exploitation.”

are often provided and very valuable in their own right, but do not always help much in getting a feeling for what is to be considered information fusion and what is not. It is sometimes tempting to paraphrase US Supreme Court Justice Stewart

Potter and say that “I can’t define what information fusion is, but I know it when I see it”. In the same sense as it may be hard to clearly separate what is information fusion from what is not, it can arguably be quite hard to get an understanding of who the “main players” in the field of information fusion are:

- Who are the most influential researchers?
- Which are the strong research groups?
- Who is collaborating with whom?
- What researchers share our research interests?

We attempt to contribute to a shared understanding of the information fusion field by doing a social network analysis of the research community. To try to understand who the main actors are is in many aspects not very different from the analysis undertaken by analysts trying to identify people of interest in criminal or terrorist networks (often referred to as *dark networks* [3], [4]). As an example, [5] discusses what type of network data that is used by detectives when deciding whom to bring in for questioning in criminal investigations (e.g. node centrality, betweenness centrality, subgroup identification, and structural equivalence). Another example is the capture of Saddam Hussein, which was a result of a social network analysis where the tribal and family linkages of Hussein were traced, in which certain individuals who may have had close ties to him were identified (see [6], [7]).

In this paper, we use bibliometric tools and methods as well as social network analysis in order to create an overview of the information fusion research field. The purpose of the overview is twofold, in that it provides an illustration of how methods from social network analysis can be used to support intelligence analysis (on data that is familiar to people in the information fusion community and thereby hopefully supports a better understanding of the concepts), but also provides a view of the field of information fusion as such. This overview is created by making use of structured information collected from Thomson Reuters’ ISI Web of Knowledge¹ and IEEE Xplore².

The rest of this paper is structured as follows. In Section II, we provide the reader with a background to social network analysis and give a brief overview of related work. In Section III, we describe the datasets that have been used and how

¹www.isiknowledge.com

²www.ieee.org

they have been collected. We also present results from various analyses that have been made, such as which authors that have contributed to most publications in the fusion domain, what words that are most commonly used in articles related to information fusion, and which authors that are most central according to various social network analysis measures. Section IV provides a discussion of the major differences between constructing social networks based on co-authorship (as has been the case in this work) and less well-structured data sources with a higher degree of imperfect information, more likely to be the case in e.g. the intelligence analysis domain. We also compare the results of our analyses to those obtained for other research communities. Finally, we conclude the paper in Section V and present thoughts for future work.

II. BACKGROUND

A. Social network analysis

To many people, the notion of a social network is equivalent to social networking services on the Internet such as Facebook and LinkedIn. However, the scope of social networks is much broader than that. Research on social network analysis (SNA) stretches back more than half a century [8], [9], where Jacob L. Moreno often is credited to be the researcher who was first to systematically make use of social network analysis-like techniques [10] (for a more complete overview of the history of social network analysis, see [11]).

A social network is often represented as a graph $\mathcal{G} = (\mathbf{V}, \mathbf{E})$ consisting of a set of nodes \mathbf{V} and a set of edges \mathbf{E} , where the nodes (vertices) typically are used to represent actors in the networks (e.g. persons, teams, or organizations), while the edges (also often referred to as ties in the terminology relevant to social networks) represent relationships among actors (such as kinship, communication, business relationship, etc.). The edges can be either directed or undirected, depending on the type of network that is modeled. An example of a social network is illustrated in Figure 1, in which the nodes represent people, and where a directed edge from A to B can be interpreted as indicating that person A has made a phone call to person B during some specified period of time. In this case, $\mathbf{V} = \{Alice, Bob, Claire, David, Eve\}$, while $\mathbf{E} = \{(Alice, Claire), (Bob, Alice), (Bob, David), (Bob, Eve), (Eve, David)\}$. In this very limited social network, we can observe differences among nodes such as Bob having many outgoing phone calls but no incoming, David being the one having most incoming calls, etc.

Social network analysis is of interest in a vast amount of areas. To mention just a few, it can be used for understanding social interactions, to optimize flow of information between employees in a company, or to study and analyze criminal or terrorist organizations. Important problems within social network analysis are, among others:

- 1) to collect and extract useful data,
- 2) to visualize the network in a way that support analysts with interpretation of the social structures, and

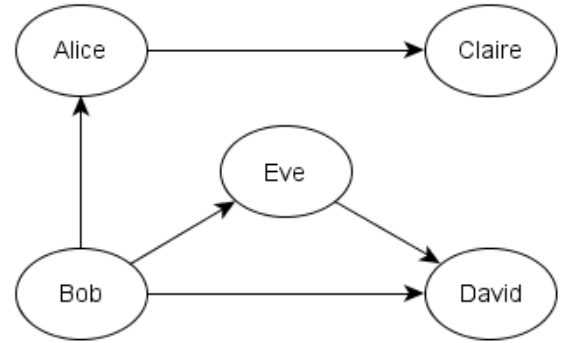


Figure 1. An example of a small social network.

- 3) to identify important structural patterns of the network, (such as the identification of actors in the network that are extra important or powerful).

All of these problems are highly relevant to this paper, although our main focus is on the last one.

Once a graph has been constructed in one way or another (e.g. through interviews, questionnaires, direct observation, data from archival records, etc.), there are a number of measures that can be used to determine the importance of a node in the network. Two such important measures are *degree centrality* and *betweenness centrality*. The degree centrality of a node is simply the number of edges that the node has, while betweenness centrality measures the proportion of *geodesics* [8], i.e. shortest paths between two nodes in the network, that flow through the particular node. The larger the number of shortest paths that passes through a node, the more central to the network it can be argued to be. Moreover, removing nodes with maximum betweenness from the network will typically result in large increases in minimum distances among the other nodes [12]. The nodes with high betweenness are often referred to as *information brokers* or *gatekeepers* [8].

In addition to measuring the impact of individual actors in the network, it can be of importance to analyze properties of the network as a whole. *Centralization* is one such property, measuring the relative difference between the highest and lowest values for the betweenness centrality measure over all nodes in the graph. Hence, this measure gives an idea of the variability of centrality among nodes in the network [9]. Another widely used social network analysis concept is that of *density*, defined as the ratio of the number of edges in the graph and the maximum possible number of edges in a graph with the same number of nodes [8], [9]. Hence, the lowest possible density is 0 (no edges present in the graph) and the highest possible density is 1 (a complete/fully connected graph).

Studies of research communities using bibliometric methods have previously been undertaken for several different research fields. To our knowledge, however, this is the first one for the information fusion community. A famous example of a social network based on co-authorship is that of the Hungarian mathematician Paul Erdős. Since Erdős published extremely

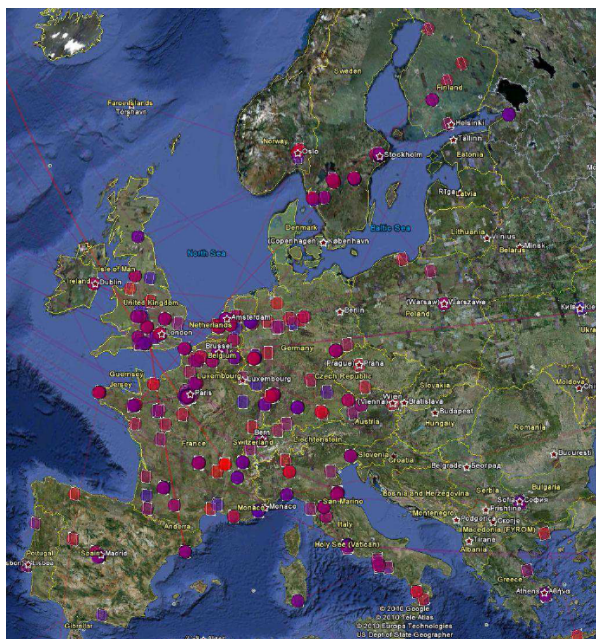


Figure 3. Geographical distribution of the FUSION community in Europe.

Fusion during the period stated above. Among these, Éloi Bossé is in the top of list with four journal articles, as can be seen in Table I. He is closely followed by Stavri Nikolov, Rick Blum, John Lewis, and Dave Bull that have contributed with three articles each. 31 researchers have been authoring or co-authoring two articles, and another 323 researchers have contributed to one article in *Information Fusion* during that period of time.

C. Analysis of dataset 2: Information Fusion conference

A more detailed analysis has been conducted using the larger dataset from Web of Science, i.e. dataset D_2 . This has been used as input to CiteSpace (see [23]), in which we have created a .kmz-file (a compressed keyhole markup language-file, which is the format used by e.g. Google Earth). In this file, the authors' address information have been mapped to geographical coordinates, and links have been created between the locations of co-authors (the reason why this was not done for D_3 instead is that address information is not part of the records downloaded from IEEE Xplore). In this way, a geographical mapping of major parts of the information fusion community has been created. Figure 3⁵ gives an idea of how the information fusion community is geographically distributed in Europe.

Based on the authors' address information we can also see which countries and continents that are the most prominent in the fusion community. Figure 4 shows the proportions of the number of articles published by authors from different countries, and in Figure 5 the same information has been

⁵This and all other figures will be available on www.foi.se/fusion in high-resolution and with zoom possibility.

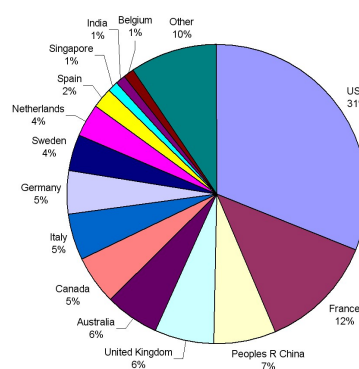


Figure 4. Proportion of articles from various countries.

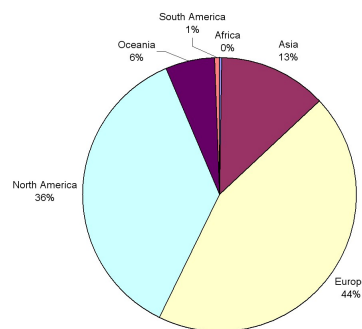


Figure 5. Proportion of articles from various continents.

aggregated into continents. The dominance of Europe and North America is perhaps not that unexpected, but it is remarkable how low presence South America and Africa have in the community.

Dataset D_2 has also been used in order to identify the main organizations contributing to the information fusion field. When doing the analysis we discovered that too many variants of organization names have been used in various papers to allow for a completely fair comparison. As an example, there were seven hits for *Def R&D Canada Valcartier*, two for *Def R&D Canada*, one for *Def Res & Dev Canada*, etc. We manually aggregated names that seemed to originate from the same organization, but some errors are still likely to exist in the used dataset. With this said, some of the most "influential" organizations and institutions (based solely on the number of papers) that were found in our analysis are: the Defence Science and Technology Organisation (DSTO), University of Melbourne, University of New Orleans, Xian Jiaotong University, US Air Force⁶, Defence Research and Development Canada (DRDC), and the Swedish Defence Research Agency (FOI), all with more than 20 papers each.

D. Analysis of dataset 3

By processing the largest dataset consisting of 2412 records in Bibexcel, we noticed that X. Rong Li is in the top with

⁶This is an aggregate of e.g. Air Force Research Laboratory and Air Force Institute of Technology.

astounding 55 papers, closely followed by Chongzhao Han with 50 papers. The top five authors (measured in number of papers) are shown in Table II. It should be noted that there is some issues with variants of names also here. Some small differences among the used data formats in how they represent names have been taken care of, but there are also problems that are harder to handle. As an example, our FOI colleague Ronnie Johansson is on six occasions referred to as Johansson R, two as Johansson LRM, and once as two separate persons (Ronnie L and Johansson M). Similarly, there are people in the dataset that share the same name and therefore are treated as one and the same person. One such example that has been found is Chen H, which looks as one person but actually is an aggregation of (at least) two distinct individuals: Hongda Chen and Huimin Chen. Since problems like these are very time consuming to handle and demand a lot of context knowledge, these problems have been ignored, except for the authors in the top of the lists. This again reminds us of a very important problem that intelligence analysts face, i.e. *entity resolution*.

Table II
TOP FIVE AUTHORS FROM DATASET 3 (MEASURED IN NUMBER OF PUBLICATIONS)

Author	Number of papers
Li XR	55
Han CZ	50
Blasch E	46
Ristic B	38
Hanebeck UD	33

In the next step we have constructed a co-occurrence matrix in Bibexcel, and from this matrix created a social network consisting of all the authors identified in dataset D_3 . This network (from now on referred to as the FUSION network) consists of 3305 nodes (authors) and 7203 edges (where two nodes are connected if their corresponding authors have written a publication together), and have been transformed to a .net-file readable by the software Pajek [24]. A problem with social networks containing a large number of nodes and edges is that they easily become hard to overlook. For this reason, we have converted the network to a GraphML-file readable by the visualization toolkit Prefuse⁷. In this way, we have written Java code that makes it possible to interact with the network by allowing for moving around nodes, and to zoom in and out on interesting parts of the network. This kind of interaction allows for a much better understanding and analysis of the network. For illustration purposes, an overview of the FUSION network is shown in Figure 6, while a very small zoomed in part of the network is shown in Figure 7 (where the numbers after the \$-signs correspond to the number of publications of each author).

E. Applying SNA concepts on the constructed social network

The FUSION network has been further analyzed by applying various social network analysis measures such as degree

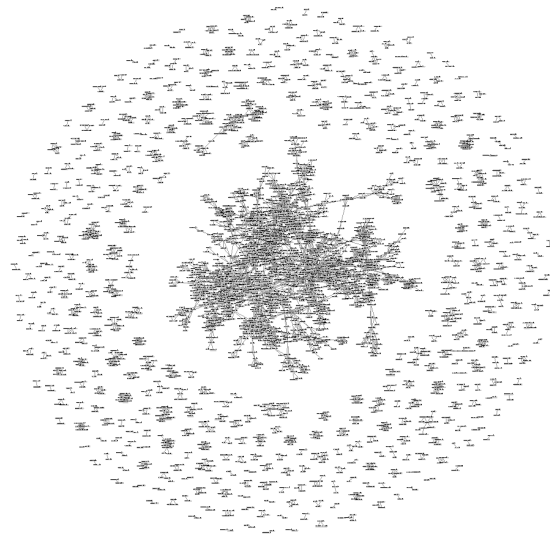


Figure 6. An overview of the FUSION co-author social network.

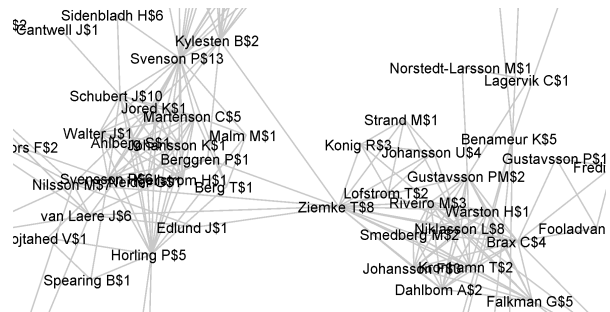


Figure 7. Zoomed in part of the FUSION co-author social network.

centrality, betweenness centrality, and density. The created .net-file has been imported into UCINET⁸ (see [25], [26]), which has been used together with Pajek to undertake such analyses. As discussed briefly in Section II-A, a node's degree centrality is simply its number of connections. When analyzing the degree centrality of the co-authorship network, the researcher who is most well-connected in the fusion community is Erik Blasch. As we can see in Table III, he has a degree of 72, while the average degree in the network is 4.4. As also was described in Section II-A, a node's betweenness centrality is the number of geodesic paths that passes through it. Analysis of this property reveals that Yaakov Bar-Shalom is the researcher with highest betweenness centrality in the network, closely followed by Erik Blasch (see Table IV).

The density of the network is approximately 0.0013. Hence, the proportion of all possible co-authorships that are actually present in the network is very low (recall that the density is a number between 0 and 1). We have also calculated a centralization value for the social network as a whole, yielding a value of 0.16%. Collaboration networks with high

⁷<http://www.prefuse.org/>

⁸www.analytictech.com/ucinet/

Table III
TOP FIVE AUTHORS FROM DATASET 3 (BASED ON DEGREE CENTRALITY)

Author	Degree centrality
Blasch E	72
Hanebeck UD	67
Li XR	60
Willett P	58
Farina A	57

Table IV
TOP FIVE AUTHORS FROM DATASET 3 (BASED ON BETWEENNESS CENTRALITY)

Author	Betweenness centrality
Bar-Shalom Y	7.2
Blasch E	7.1
Maskell S	4.4
Kirubarajan T	4.1
Chen H	4.1

centralization can be said to consist of a few highly dominant researchers, while lower values indicate a research community in which contributing authors are more equal in their centrality scores [14]. The most centralized network topology possible is the star topology, and the centralization result should be interpreted as the degree of inequality in the FUSION network as a percentage of that of a perfect star network of the same size.

Moreover, we have calculated the average distance in the network to be approximately 7.0. It should however be noted that the FUSION network actually consists of one large component (often called the connected core in network science) and several very small components which are disjoint from the large component (see Figure 6). Hence, when calculating the average distance, this has been calculated among reachable pairs. A summary of statistics for the information fusion community is given in Table V.

Table V
STATISTICS FOR THE FUSION COMMUNITY (YEARS 2002–2010)

Tot. nr of papers	2412
Tot. nr of authors	3305
Mean nr of authors per paper	2.9
Mean nr of collaborators per author	4.4
Mean distance	7.0
Centralization	0.16%
Density	0.0013

F. Community detection

As briefly mentioned in Section III-E, the FUSION network actually consists of a large connected core and a couple of smaller components which are not connected to the large component (once again, in terms of co-authorship). These can be viewed as different clusters, but obviously there are also various sub-communities within the large component. To identify subgroups in e.g. terrorist networks can be of very high relevance for intelligence analysts, and we will here demonstrate this task by identifying sub-communities

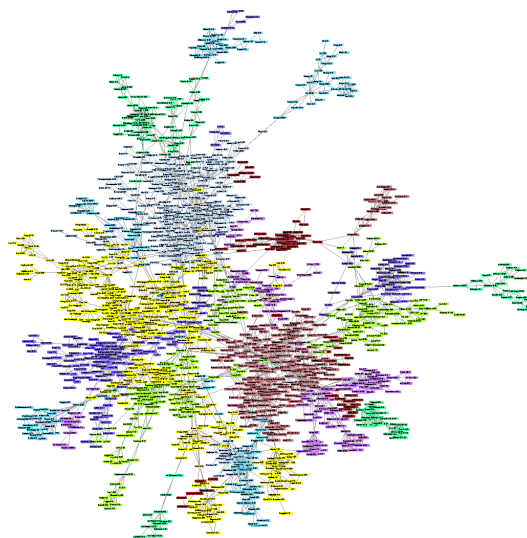


Figure 8. An overview of various sub-communities within the main component of the FUSION network.

in the FUSION network. The topic of *community detection* has in recent years been a popular research area and a good introduction is given in [27]. The algorithm that has been used for community detection in our experiments is the so-called *spin glass method*. The name of algorithm stems from an analogy to how community detection can be interpreted as an optimization problem where we search for the ground state of an infinite range Potts spin glass. For a detailed description of how the algorithm works, we refer the interested reader to [28].

When the spin glass algorithm was applied to the main component of the FUSION network without any constraints, 45 sub-clusters were identified. Since this is too many sub-clusters to actually provide any useful insight, we have limited the number of clusters to 10 before the algorithm was run a second time. The resulting sub-clusters from the second run of the algorithm are visualized using different colors in Figure 8. To judge the overall quality of these sub-clusters is quite hard, but a visual inspection of the created clusters suggests that most of the created clusters are reasonable. As an example, if we look at various Swedish researchers, these tend to end up in one of two clusters: one cluster involving people from University of Skövde, and one cluster involving people from Swedish Defence Research Agency (FOI).

IV. DISCUSSION

Table VI
COMPARISON OF COMMUNITY STATISTICS

Community	Mean dist.	Coll. per auth.	Auth. per paper
FUSION	7.0	4.4	2.9
MEDLINE	4.6	18.1	3.8
SPIRES	4.0	173	9.0
NCSTRL	9.7	3.6	2.2

The average distance in the FUSION network can be compared to that of MEDLINE (biomedical research), SPIRES (high-energy physics), and NCSRTL (computer science) reported in [16], as shown in Table VI. One possible interpretation of these numbers is that the information fusion community is not as tight as e.g. that of high-energy physics. An explanation to this can be that information fusion is a mix of various disciplines, but an alternative explanation can also be that this is due to the focus on defense applications which sometimes can make it hard for researchers to cooperate over organization or country boundaries. We can in the same way compare the average number of collaborators per author and number of authors per paper in the various research fields. As seen in Table VI, the SPIRES network clearly is an extreme outlier compared to the other research fields (explained with many mega-collaborations within high-energy physics), but it is as interesting to see that there are more collaborations between researchers in the field of information fusion than within the somewhat more theoretically oriented field of computer science.

In the study presented in this paper, the actual construction of the co-authorship social network was quite straightforward due to the well-structured information downloaded from Web of Science and IEEE Xplore. This can be compared to what the construction phase would look like if the social network to analyze instead would have been a dark network. For such a problem, the construction phase of a graph representing the social network would demand considerably more work, since the data from which such a graph would be constructed is likely to be less well-structured and harder to collect. Moreover, the data for a dark network is also likely to be imperfect to a higher degree than has been the case with the co-authorship data. As an example, if the social network of a terrorist group was to be constructed from intelligence reports, there would be uncertainty of whether there actually should be a link between two members of the network or not. This is not the case with the social network we have studied here, either two researchers have co-authored a paper or they have not (obviously there can be other kinds of relationships between two researchers than just the co-authoring of papers, but here we only care about this type of social connection). We can of course choose to model only relationships that we are certain exists also for dark networks, but it is likely that we then would miss valuable information. However, if we choose to link together all pairs of criminals for which we have some kind of intelligence suggesting that there might be a possible connection, we are instead likely to come up with a densely connected network in which it is more connections than actually should be there. A natural extension is therefore to add a weight to each edge in the network, where each edge says something about the probability or strength of the relationship.

As has been described in Section III, there were some problems with non-unique author names in the downloaded records. The same author can be referred to in slightly different ways in various records (e.g. due to misspelling of names,

inclusion/exclusion of initials, or name changes). Likewise, it is possible that two individuals share the same name, or have names that are very similar to each other. Similar kinds of problems are likely to arise also when analyzing dark networks. A person in the network may use several aliases, making it possible that two nodes in the network actually should be merged together into one and the same node. There may also be alternative spellings of a name. Are Moamar El Kadhaafi and Moammar Gadhafi referring to the same individual or not? Such problems are non-trivial to manage. One way to attack the problem of determining whether two nodes are referring to the same individual or not in the context of the FUSION co-author network studied in this paper is to compare whether they belong to the same university/organization or not (for the records where such information exists), and whether there is some overlap between the nodes they are connected to. It is also possible to include other sources (e.g. to download papers and control whether the full name of two authors match or not, or by checking the conference programs). Even though such strategies can be used to eliminate some problems, very much context knowledge is demanded to keep track of changes such as people getting married and changing their surname. Another strategy for managing some of the uncertainty is to construct two networks, one where we use surname and first initial only to identify individuals, and one where a stricter definition of what constitutes a unique individual (e.g. full name together with organization information) is used. In this way, we tend to underestimate the number of individuals in the first case, and to overestimate it in the second case. Hence, this can be used as lower and upper bounds on the number of authors (as well as many of the used social network analysis measures). An example of where such an approach is used is [17], however, this approach has not been used here since we do not have full name or organizational affiliation available for all records.

V. CONCLUSIONS

We presented the fundamentals of social network analysis. By collecting bibliographic records of which authors that have been publishing in proceedings of the international conference on information fusion, and the *Information Fusion* journal, we have constructed a social network for the information fusion domain, in which nodes in the network represent authors, and where edges in the network represent co-authorship. Various social network analysis measures such as degree centrality have been applied to this network in order to identify the most “powerful” scientists in the domain. By most powerful we are not only referring to authors that have produced most publications, but also to authors which are connected to most other authors, and authors that work as bridges between sub-clusters that otherwise would have been isolated. We have also provided a coarse overview of where the white spots of the information fusion community are geographically speaking, and which words that are commonly used within the titles of information fusion-related articles. The provided social network analysis of the information fusion community is

thought to be valuable in its own right, but the purpose of the paper is also to show the workings of social network analysis concepts on a domain that is well-known to many of us, so that we become more prepared to use this kind of techniques on e.g. dark networks, for which highlighted problems such as collection of data and determining whether two nodes are referring to the same individual or not become even harder.

A. Future work

An important problem that has become evident when working with the data is that we need to be able to work with uncertainties if the same kind of techniques that have been used in this paper are supposed to be used also for relationships that are not as clear cut as the co-authorship between two authors. To make use of weights for indicating the strength or probability of a relationship being present is nothing new, but more work is needed on how to establish such weights based on e.g. intelligence reports associated with an information content rating and a reliability of the source. Some preliminary ideas for how this could be done using simulations of uncertain networks has earlier been presented in [29]. We are currently investigating how community detection can work on uncertain networks. In this work, the uncertainty is handled by simulation; a large number of samples from an ensemble of networks that are consistent with the known information are used to determine several different community structures. These community structures are then fused using ensemble clustering methods.

Although the analysis in this paper gives some hints on who the main players in the fusion community might be, a different take would be to study the citation network of the community to identify influential papers and authors. The citation network could also be used to study what other conferences and journals that are important to fusion researchers. Another aspect of interest would be to study how the community has evolved over time. One possibility is to compare the social networks from different time periods, but one could also study how the content evolves by comparing tag clouds.

For the future we also would like to see JAIF articles to be part of Web of Science and IEEEExplore.

ACKNOWLEDGEMENT

We would like to thank Johan Dahlin for providing us with the code for community detection.

This work was supported by the FOI research project "Tools for information management and analysis", which is funded by the R&D programme of the Swedish Armed Forces.

REFERENCES

- [1] L. Valet, G. Mauris, and P. Bolon, "A statistical overview of recent literature in information fusion," in *Proceedings of the Third International Conference on Information Fusion*, 2000.
- [2] B. V. Dasarathy, "Information fusion - what, where, why, and how?" *Information Fusion*, vol. 2, pp. 75–76, 2001.
- [3] J. Raab and H. B. Milward, "Dark networks as problems," *Journal of Public Administration Research and Theory*, vol. 13, pp. 413–439, 2003.

- [4] P. Svenson, P. Svensson, and H. Tullberg, "Social network analysis and information fusion for anti-terrorism," in *Proceedings of the Conference on Civil and Military Readiness 2006*, 2006.
- [5] B. Zhu, S. Watts, and H. Chen, "Visualizing social network concepts," *Decision Support Systems*, vol. 49, no. 2, pp. 151–161, 2010.
- [6] V. Hougham. (2005) Sociological skills used in the capture of Saddam Hussein. <http://www.asanet.org/footnotes/julyaugust05/fn3.html>.
- [7] D. Koelle, J. Pfautz, M. Farry, Z. Cox, G. Catto, and J. Campolongo, "Applications of bayesian belief networks in social network analysis," in *Proceedings of the 4th Bayesian Modeling Applications Workshop during the 22nd Annual Conference on Uncertainty in Artificial Intelligence*, 2006.
- [8] J. Scott, *Social Network Analysis*, 2nd edition, Ed. SAGE Publications, 2000.
- [9] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.
- [10] M. E. J. Newman, *Networks: An Introduction*. Oxford University Press, 2010.
- [11] L. C. Freeman, *The Development of Social Network Analysis: A Study in the Sociology of Science*. Empirical Press, 2004.
- [12] M. E. J. Newman, "Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality," *Physical Review E*, vol. 64, pp. 016 132–1–016 132–7, 2001.
- [13] M. Biryukov and C. Dong, *Lecture Notes in Computer Science*. Springer-Verlag, 2010, vol. 6273/, ch. Analysis of Computer Science Communities Based on DBLP, pp. 228–235.
- [14] C. Bird, E. Barr, A. Nash, P. Devanbu, V. Filkov, and Z. Su, "Structure and dynamics of research collaboration in computer science," in *Proceedings of the Ninth SIAM International Conference on Data Mining*, 2009.
- [15] M. E. J. Newman, "The structure of scientific collaboration networks," in *Proceedings of the National Academy of Sciences of the United States of America*, 2001.
- [16] M. E. Newman, *Lecture Notes in Physics*. Springer, 2004, vol. 650, ch. Who Is the Best Connected Scientist? A Study of Scientific Coauthorship Networks, pp. 337–370.
- [17] M. E. J. Newman, "Scientific collaboration networks. i. network construction and fundamental results," *Physical Review E*, vol. 64, pp. 016 131–1–016 131–8, 2001.
- [18] P. Chen and S. Redner, "Community structure of the physical review citation network," *Journal of Informetrics*, vol. 4, pp. 278–290, 2010.
- [19] S. Redner, "How popular is your paper? an empirical study of the citation distribution," *The European Physical Journal B*, vol. 4, pp. 131–134, 1998.
- [20] A. Goldenberg and A. W. Moore, "Bayes net graphs to understand co-authorship networks?" in *Proceedings of the 3rd international workshop on Link discovery*, 2005.
- [21] S. Redner, "Citation statistics from 110 years of physical review," *Physics Today*, vol. 58, pp. 49–54, 2005.
- [22] F. Åström, R. Danell, B. Larsen, and J. Wiborg Schneider, in *Celebrating scholarly communication studies: A Festschrift for Olle Persson at his 60th Birthday*. Umeå University, 2009.
- [23] C. Chen, "Searching for intellectual turning points: Progressive knowledge domain visualization," in *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 2004.
- [24] V. Batagelj and A. Mrvar, "Pajek – analysis and visualization of large networks," in *Graph Drawing*, ser. Lecture Notes in Computer Science, P. Mutzel, M. Jünger, and S. Leipert, Eds. Springer Berlin / Heidelberg, 2002, vol. 2265, pp. 8–11.
- [25] R. A. Hanneman and M. Riddle, *Introduction to Social Network Methods*. University of California, Riverside, 2005.
- [26] S. P. Borgatti, M. G. Everett, and L. C. Freeman. (2002) UCINET 6 For Windows: Software for Social Network Analysis. Analytic Technologies, Harvard, MA. [Online]. Available: <http://www.analytictech.com/>
- [27] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, pp. 026 113–1–026 113–15, 2004.
- [28] J. Reichardt and S. Bornholdt, "Statistical mechanics of community detection," *Physical Review E*, vol. 74, 2006.
- [29] P. Svenson, "Social network analysis of uncertain networks," in *Proceedings of the 2nd Skövde Workshop on Information Fusion Topics*, 2008.