Conceptual Structure of Fraud Research and Its Dynamics
Full Paper

Sonia Gantman
Providence College
sgantman@providence.edu

Dmitry Zinoviev
Suffolk University
dzinoviev@suffolk.edu

Abstract

The concept of fraud, its antecedents and outcomes as well as its detection and prevention, have been discussed by both academics and practitioners for decades. The scope and complexity of the concept of fraud attracts scholars from diverse disciplines.

The purpose of our study is to gain a broader understanding of how fraud is viewed in the compendium of academic literature. Using semantic network analysis method, we explore the structure of fraud-related research and analyze the internal connections among the current areas of interest for fraud researchers. We are particularly interested in the relationship between the mainstream financial accounting and audit research and the field of information systems and technology.

Our work makes a big stride toward the understanding of current state of fraud-related research. The interdisciplinary semantic map of keywords and subject terms helps understand the trends in fraud scholarship, identify gaps and propose directions for future research.

Keywords

Semantic network analysis, semantic map, fraud, literature analysis, EBSCO

Introduction

The concept of fraud, its antecedents and outcomes as well as its detection and prevention, have been discussed in academic and practitioner oriented literature for decades. The major scandals of the turn of the 21st century drew attention of the public to accounting and corporate fraud, and thus presented new challenges for management and for the accounting profession. The passage of the Sarbanes Oxley Act in 2002 led to substantial changes in auditing practice, which gave impetus to increased interest in fraud-related issues in accounting academic literature (Hogan et al., 2008).

Authors reflecting on the growing body of academic literature on fraud invariably call for multidisciplinary approach; combining insights from business studies, behavioral sciences, criminology, law and ethics. (Ramamoorti, 2008; Trompeter et al., 2014; Cooper, 2013). Remarkably, none of these authors mention information systems as a research field capable of contributing to our understanding of various aspects of fraud, its prevention and detection.

The purpose of our study is to gain a broader understanding of how fraud is viewed in the compendium of academic literature. We explore the structure of fraud-related research and analyze the internal connections among the current areas of interest for fraud researchers. We are particularly interested in the relationship between the mainstream financial accounting and audit research and the field of information systems and technology.

We have performed semantic network analyses of twenty years of academic publications hosted by EBSCO databases and devoted to various aspects of fraud. The results of the first stage of the project are reported in this paper.

Our work makes a big stride toward the understanding of current interests in the research community regarding fraud issues. The interdisciplinary semantic map of keywords and subject terms helps understand the trends in fraud scholarship, identify gaps and propose directions for future research.

The remainder of the paper is structured as follows. In the next section, we debrief reflections on the current state of fraud research by accounting and auditing scholars. We further introduce our research method - semantic network analysis. The data collection procedures and characteristics of the resulting data set are reported next, followed by presentation of findings. The paper concludes with discussion and plans for the next stages of the project.

Background

The phenomenon of fraud, and white-collar crime in general, is not a new topic in academic literature. In the 1940s and 1950s, sociologist E. Sutherland introduced the notion of white-collar crime (as opposed to street crime), and criminologist D. Cressey developed a theoretical model commonly referred to as "Fraud Triangle" (in Dorminey et al., 2012). During the subsequent five decades, research on fraud continued (e.g. Darby and Karni, 1973; Beasley 1996, Fawcett and Provost 1997), however it was not until the 2000s that extensions to the Fraud Triangle theory started to emerge (e.g. Albrecht et al. 2006).

The scandals of the turn of the 21st century exposed corporate fraud of unprecedented scale. They drew substantial attention of the public to the antecedents and consequences of financial statement fraud, and presented new challenges to corporate management and to the accounting profession. The Sarbanes Oxley Act of 2002 extensively addressed measures for fraud prevention and detection, as well as penalties for committing fraud; the Public Company Accounting Oversight Board (PCAOB) included financial statement fraud in its top priorities in standard-setting (PCAOB, 2004). At the same time, fraud-related issues garnered increased attention of the academic community, predominantly in the accounting and auditing fields (Hogan et al., 2008, Trompeter et al., 2013).

Dorminey et al. (2012) suggested expanding the Fraud Triangle, which is exclusively focused on the motivation for fraud, to a more complex model which includes the components of fraud, the probability of its occurrence and the role of corporate governance and internal control in fraud prevention. Along with other authors (e.g. Ramamoorti, 2008, Brody et al., 2012), they argue that the behavioral nature of fraud calls for interdisciplinary approach of accounting and social sciences scholars.

Brody et al. (2012) summarize a panel discussion on emerging issues of fraud research and express a concern with a somewhat narrow perception of fraud and fraudsters among the accounting professionals as well as the general public and the media. Brody et al. (2012) discuss assumptions about fraudsters' behavior that are not grounded in research but prevail among professionals and researchers, and call for a close collaboration with behavioral scientists.

A synthesis of a large body of fraud-related academic research by Trompeter et al. (2013, 2014) draws on the meta-model of Dorminey et al. (2012) and includes, in addition to accounting and auditing publications, journal articles from criminology, ethics, finance, organizational behavior, psychology, and sociology. Some relevant fields, however, remain omitted from their attention – thus, the two extensive bibliographies do not include a single publication from information systems or accounting information systems literature.

Finally, Cooper et al. (2013) offer a broader insight into the notion of fraud and into the context where it should be placed. They call for redefining the concept of fraud and expanding the scope of fraud-focused research to the organizational and societal levels. Along with other gaps in the fraud scholarship, Cooper et al. (2013) identify a significant gap in the understanding of how the actual fraudulent act is carried out, in their words, "the failure to examine the technologies of fraud". While "technologies" in this quote does not refer to information technology, studying details of fraudulent activity presents a clear opportunity for a collaboration of accounting and information systems scholars.

Fraud is a complex phenomenon that can be studied from different perspectives, and it allows for the application of theories and approaches from a number of research fields. It appears that current academic research on fraud and fraud-related issues is rather fragmented and leaves out many important aspects and details. In response to the call of Cooper et al. (2013) to re-conceptualize the notion of fraud and to broaden the context for its analysis, we take the first step toward this goal by exploring the current literature on fraud and analyzing the contexts for studying fraud in existing academic publications.

The first stage of our project aims at identifying the main areas of interest of fraud researchers regardless of the formal research field they represent. Analyzing the internal connections, or overlaps, among these areas of interest will present us with a better perspective on the extent of fragmentation in fraud-related research. We are particularly interested in the overlap of mainstream accounting and audit research on fraud with the literature pertaining to corporate and IT governance and the role of information systems and technology in preventing and detecting fraud.

Data Collection and Analysis

Our results are based on application of semantic network analysis to metadata collected from EBSCO – Electronic Business Source Complete.

Semantic network analysis uses the natural language of the participants. The analyzed content is represented as a semantic network. The community structure of the network, once extracted, corresponds to the topics of discourse – categories of terms defined by shared meaning (Rice and Danowski, 1993). Unlike the traditional content analysis (Pérez-Corona et al. 2012; Salton et al. 1975), semantic network analysis does not require a priori organization of the categories which is subject to coders' bias and therefore has reliability and validity limitations.

Our analysis uses terms – single words and multi-word expressions (such as "fraud" and "fraud prevention"), attached as metadata to artifacts – scholarly papers hosted by EBSCO, an aggregator of full-text academic content. We selected EBSCO as our data source for its comprehensive content and wide audience. Author-defined keywords and editor-selected subject tags provided by EBSCO define the context of each academic paper for analyzing fraud. Since our study is exploratory in nature, and we are interested in perceptions of fraud-related phenomena with regards to all members of the academic community, both keywords and subject tags are included in the dataset, and will be hereafter referred to as "terms."

Our dataset is comprised of academic peer-reviewed articles published between 1995 and 2016 and is retrieved from three EBSCO collections: Academic Search Complete, Business Source Complete, and Computers & Applied Sciences Complete. The criterion for inclusion was the presence of words "fraud" or "fraudulent" anywhere in the metadata (title, abstract or tags).

The dataset includes 15,294 articles (of which 1,878 were duplicated, resulting in 13,416 distinct articles) from ca. 3,200 journals.

Table 1 shows the distribution of publications in the dataset over the selected time period. The interest to fraud-related issues in the academic literature stayed stable during 1997-2001, then experienced a rapid growth, with number of papers per year tripling from 274 in 2001 to 894 in 2009. In the recent years (starting from 2009), the number of academic papers addressing fraud does not grow, but remains high.

Year	No of papers	Year	No of papers	Year	No of papers
1995	140	**2002**	375	**2009**	894
1996	189	**2003**	437	**2010**	913
1997	275	**2004**	576	**2011**	924

1998	268	2005	576	2012	957
1999	340	2006	704	2013	888
2000	277	2007	660	2014	941
2001	274	2008	772	2015	1032
				2016	910

Table 1. Number of academic papers on fraud published yearly between 1995 and 2016.

After eliminating stop words ("a", "the", "in", etc.), we lemmatized each term or each word in a compound term with WordNet lemmatizer. The lemmatized corpus has 24,545 distinctive terms (on average 6.7 terms per paper). We intentionally excluded terms "periodical", "letter, editor", "case, study", "research", "research, article", "serial, publication", and "conference, convention", as referring to the articles genres rather than to the contents.

We observed that some terms (such as "forced, labor" and "human, trafficking" or "internal, auditing" and "internal, auditor") very frequently co-occur in metadata. To preserve the proximity, we combined them into 193 synonym sets (synsets). Hereafter, we treat synsets indistinguishably from terms. Table 2 presents 50 most frequent fraud-related terms in the corpus and their change of frequency over time.

	Term	Frequency	II vs I	III vs II
1	financial, statement | accounting, fraud	790	0.88	0.73
2	white, collar, crime | economic, crime | commercial, crime	711	1.08	0.42
3	auditor | auditing	664	0.56	0.61
4	medical, research | fraud, science | scientist	620	0.47	0.44
5	internet, fraud | computer, crime | phishing	555	1.5	0.46
6	fraud, prevention	514	0.31	1.63
7	judgment, law | action, defense, law	503	1.11	0.52
8	insider, trading, security | security, fraud | security	477	0.51	0.54
9	ethics | business, ethics	423	0.89	0.48
10	financial, institution | banking, industry | bank, fraud	381	1.08	0.64
11	political, corruption | election, corrupt, practice | election	371	0.67	0.94
12	taxation, law, legislation | tax, evasion | taxation	355	0.58	0.94
13	internal, auditing | internal, auditor	355	0.6	0.71
14	accounting	355	0.72	0.44
15	industrial, management | corporate, governance	345	1.54	0.61
16	advertising | consumer, protection | false, advertising	314	0.59	0.44
17	data, mining | outlier, detection | fraud, detection	288	3.48	1.59
18	insurance, company | insurance | insurance, crime	262	0.51	0.53
19	organizational, behavior | corruption	260	1.89	1.5
20	fraud, investigation	253	0.27	1.1
21	bankruptcy | fraudulent, conveyance | debtor, creditor	243	0.49	0.81
22	forensic, accounting	240	1.14	0.58
23	legislation | law | legislative, bill	191	1.71	0.3
24	crime	191	1.56	0.44
25	credit, card | credit, card, fraud	191	0.65	0.8

26	capitalist, financier	190	1.97	0.71
27	Internet	189	0.73	0.46
28	electronic, commerce	185	1.66	0.51
29	medicaid, fraud | medical, care, united, state | medicare, fraud	185	0.14	0.52
30	accountant	183	0.53	0.36
31	identity, theft	176	10.16	0.42
32	computer, security	175	2.13	0.42
33	government, policy	166	0.23	1.15
34	business, enterprise	159	0.98	0.47
35	management	159	0.59	0.81
36	risk, assessment	156	0.86	0.83
37	deception	153	1.79	0.72
38	medical, care | patient	152	0.59	0.57
39	risk, management, business	148	1.66	0.77
40	whistleblower | whistleblowing	145	0.58	1.05
41	consumer, fraud	145	1.62	0.49
42	history | great, britain	144	1.46	1.46
43	statistic | data, analysis	143	1.03	2.93
44	medicaid | medical, policy | medicare	139	0.21	1.12
45	decision, making	137	0.49	1.35
46	information, technology	133	1.37	0.56
47	criminal, law	132	1.47	0.31
48	swindler, swindling | ponzi, scheme	129	1.31	1.93
49	adulteration | food, adulteration, inspection	125	x	7.34
50	accounting, corrupt, practice	120	0.93	0.37

Table 2. 50 most frequent fraud-related terms and their frequencies, as well as frequency changes for the periods 1995 − 2001 vs 2002 − 2008 (I vs II) and 2002 − 2008 vs 2009 − 2016 (II vs III). Original terms in synsets are separated by vertical bars. * The term was not available before 2002.

The frequencies of certain terms were subject to substantial changes over time. To capture the trends, we measured the frequencies in 1995--2001 (as f_1), 2001--2008 (as f_2), and 2009--2016 (as f_3), and calculated the ratios f_2/f_1 and f_3/f_2 (the last two columns in Table 2). The most volatile terms can be arranged in five groups:

- Rapidly declining: fraud in science;
- Slowly declining: white collar crime, crime in general, ethics, consumer fraud, Internet fraud, fraud prevention and investigation, Medicaid fraud;
- Rising and then declining: identity theft and computer security;
- Slowly rising: data mining and statistics;
- Rapidly rising: adulteration.

We would like to emphasize that our observations reflect, not the actual fraud trends, but the interest in those trends in the academic research community, as captured by the EBSCO corpus.

The fifty most frequent terms support observations from previous literature that the fraud-related research under-employs the behavioral perspective (Dorminey et al., 2012, Ramamoorti, 2008, Brody et al., 2012), and expresses almost no interest in understanding of how the actual fraud is carried out

(Cooper et al., 2013). There are, however, signs of expanding the scope of fraud-focused research to organizational and societal level (Ibid.), manifested through the frequent and rising use of terms "organizational behavior, corruption" (item 19) and "government, policy" (item 33).

Network construction

We arranged the selected terms into a semantic network whereby terms are represented as network nodes, and two nodes are linked with an edge if the respective terms are used together in at least five articles. Most of the nodes in the resulting network are isolated (the respective terms do not co-occur frequently enough with any other terms) and thus useless for community detection. Only 496 nodes have at least one connection of various strengths. Of those, 469 nodes form a giant connected component (GCC) – the largest part of the network where each node is reachable from any other node. The GCC of the semantic network is shown in Figure 1.
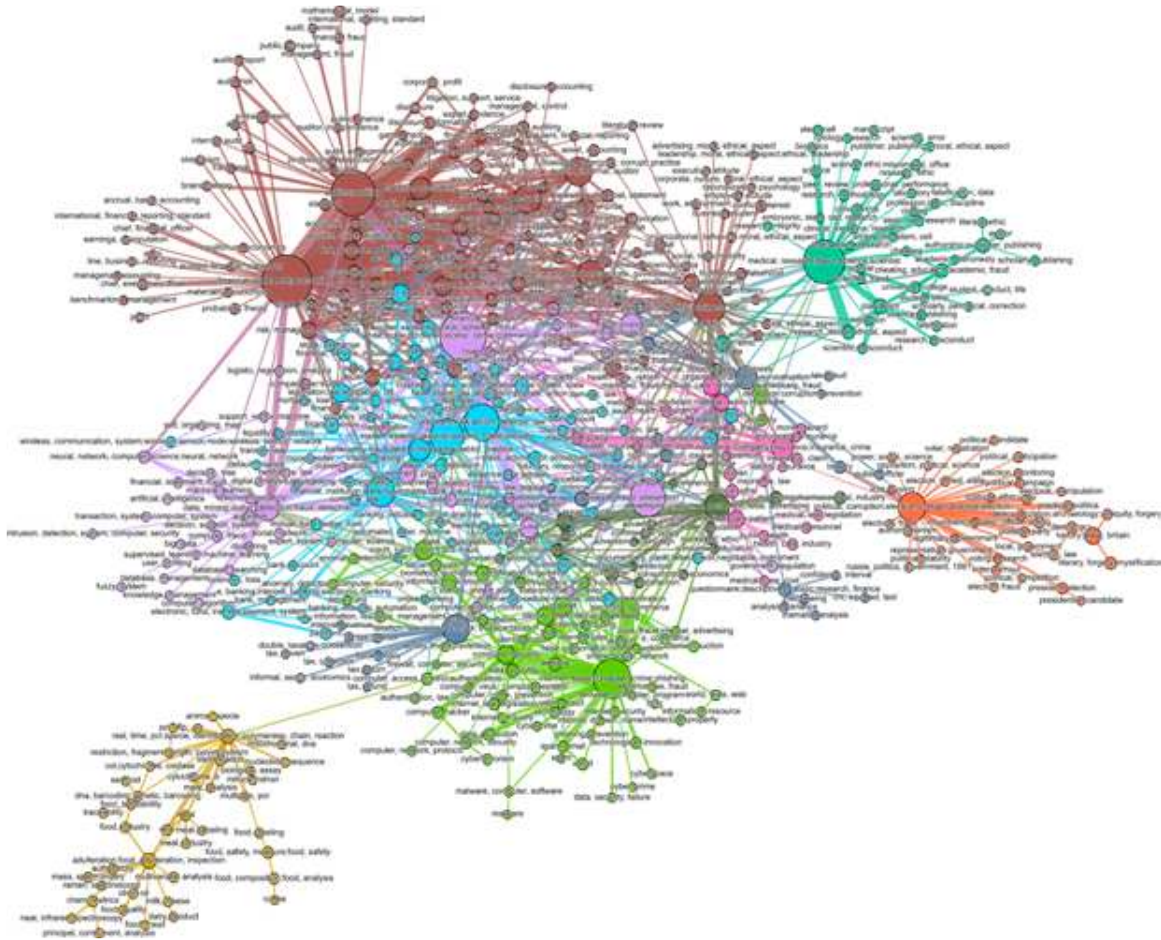

Figure 1. The giant connected component of the semantic network of 469 fraud-related terms. Node sizes represent number of articles per term.

The network of terms predictably has a high degree of modularity (Steyvers 2005): m=0.62 on the scale from -1.0 to 1.0 . The community structure is easily visible in Figure 1. We extracted communities – term clusters – using the Louvain clustering algorithm by Blondel et al. (2008). The algorithm assigns terms to clusters to minimize the number of inter-cluster links and maximize the number of intra-cluster links. We hypothesize that all terms in a well-defined cluster are substantially similar, in the sense of co-occurrence in metadata, to one another, and form a fraud-related topic. The top 50 most frequently used terms from Table 2 naturally belong to some of the topics, too. According to the community structure, there are 10 major multi-term topics in the corpus, interconnected to a various degree.

We can substantially simplify the original semantic network by creating an induced network: a network where each topic is represented as a single node of the cumulative size, and the links between the topic nodes represent the aggregated links between the individual terms (Figure 2). In the absence of algorithmic means of naming the induced nodes, we derived their names manually by inspecting the constituent terms. We focus our attention on popular and emergent topics related to accounting, audit and technology, and therefore excluded from analysis few smaller topics that address fraud in specific fields and do not overlap with the topics of our interest - election fraud, insurance fraud, fraud in science and food adulteration.
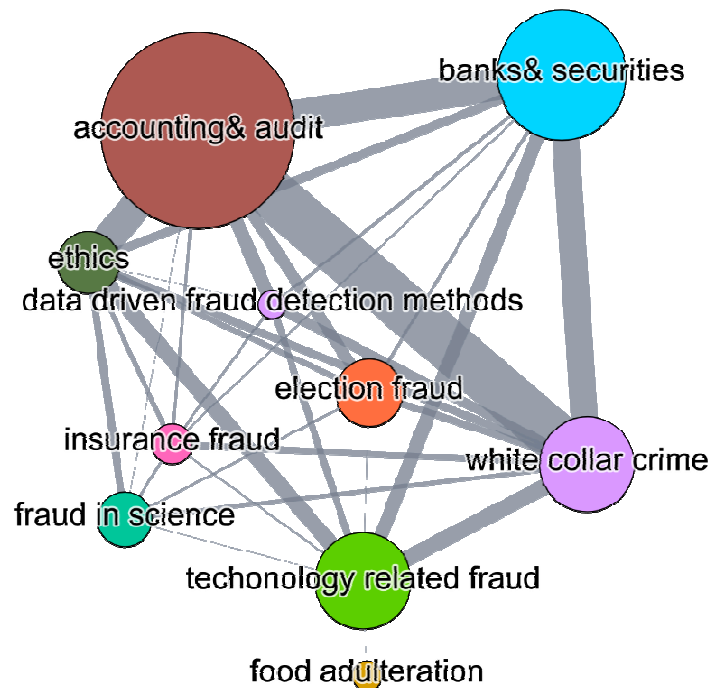

Figure 2. Induced semantic map of fraud-related topics (cf. Figure 1). Node sizes represent the number of articles per topic.

The largest nodes on the map are "accounting and audit", "banks and securities", "white collar crime" and "technology related fraud". The first three present strong connections on the semantic map, which means overlap in research in these fields.

The connection between "accounting and audit" and "ethics" nodes suggests that ethical aspects of fraud are often studied together with accounting and audit aspects. The attention to ethical aspects of fraud related to banking and securities is much weaker. Same can be said with regards to the connection between ethics scholarship and research on technology related fraud.

The lack of direct connection between the "technology related fraud" node and the "accounting and audit" node identifies that fraud researchers do not perceive these topics as related. The giant "accounting and audit" node includes mostly terms related to financial reporting and audit, with few terms related to corporate governance, internal control and risk management. The "technology related fraud" node includes such terms as "spam, email" and "internet, auction", which indeed have nothing to do with financial accounting. However, there are also many terms in the "technology related fraud" node that could appear in accounting and auditing studies. Some examples are "computer security", "data protection", "electronic commerce" and even "information technology". The lack of connection between the "accounting and auditing" node and "technology related fraud" node in the semantic map indicates a gap in fraud research that needs to be addressed.

To understand the evolution of fraud research literature, we traced the change of the network distance (the shortest path length) between some of the top 50 most frequently used terms (Table 2), namely, all terms from the topics "technology related fraud", "ethics", "banks & securities", "accounting & audit", "white collar crime", and "data driven fraud detection methods." Figures 3 and 4 show the dynamics of the closeness between the "accounting and audit" and the other four topics and "technology related fraud" and the other four topics respectively.
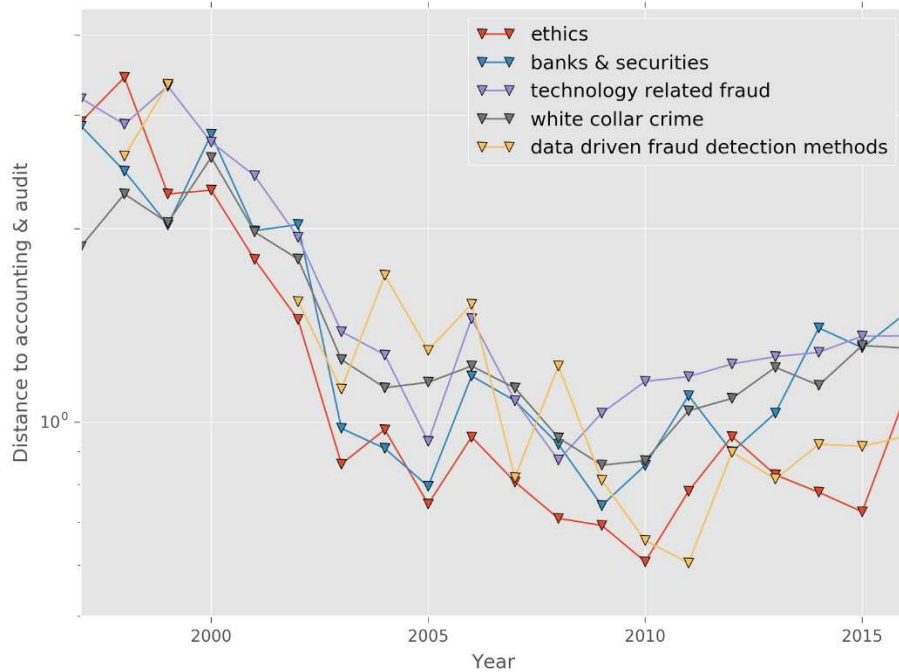

Figure 3. Semantic distance between the most prominent Accounting & Audit fraud-related topics and other select topics (cf. Figure 2).
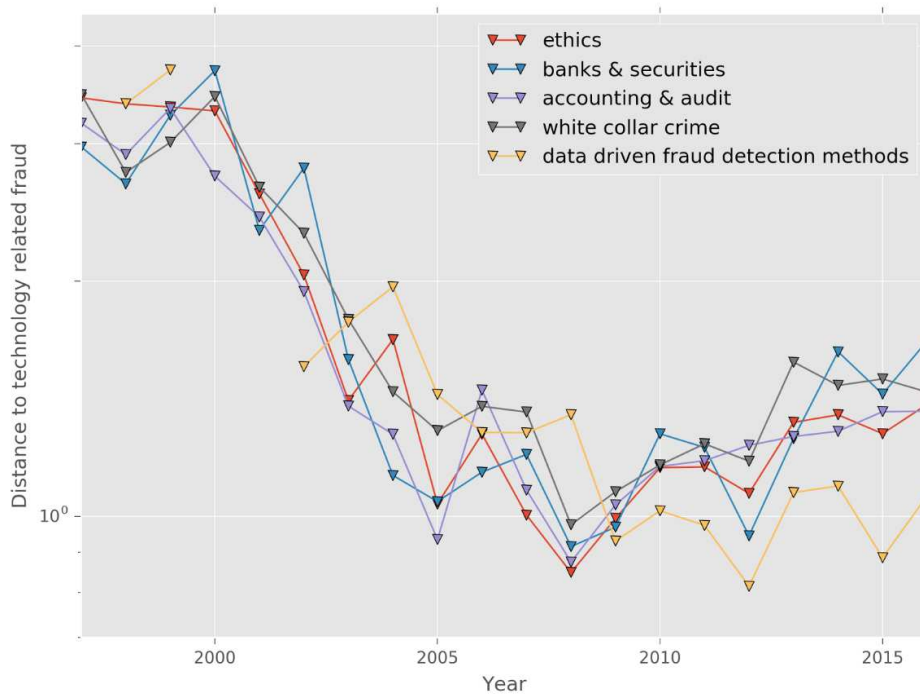
Figure 4. Semantic distance between the most prominent technology related fraud topics and other select topics (cf. Figure 2)

The semantic map and the semantic closeness graphs show the emergence of a new topic in fraud research - fraud detection methods based on big data algorithms. The smallest node on the semantic map, "data driven fraud detection methods" is the only node that becomes closer over time to both the "technology related fraud" node and the "accounting and audit" node. Growing interest to data driven fraud detection techniques is a new and promising development in approach to fraud.

Other nodes ("white collar crime", including fraud prevention, and "banks and securities), on the contrary, overlap less with "technology related fraud" and "accounting and audit" in the last five years. While showing similar patterns in relation to other topics, "accounting and audit" and "technology related fraud" remain among the most distant nodes on the map.

Discussion and Directions for future research

We presented the initial results of our research project, aimed at understanding the structure and temporal dynamics of fraud-related research across various disciplines, with a special attention to the relation between the accounting and audit field and the field of information systems and technology. We analyzed 13,416 distinct articles from over 3,200 academic journals representing a variety of business and science disciplines, built a semantic network (Figure 1), and converted it into semantic map which presents each substantial topic of fraud research with a single node and also shows the links between the topic nodes (Figure 2).

The fifty most frequent terms support observations from previous literature that the fraud-related research under-employs the behavioral perspective (Dorminey et al., 2012, Ramamoorti, 2008, Brody et al., 2012), and shows very little interest in understanding the technical details of fraudulent activities (Cooper et al., 2013). There are, however, signs of expanding the scope of fraud-focused research to organizational and societal level (Ibid.), manifested through the frequent and rising use of terms "organizational behavior, corruption" (item 19) and "government, policy" (item 33).

The semantic map of fraud research shows a strong dominance of accounting and auditing context in the fraud related academic literature, followed by the overlapping topic of banking and securities fraud. The semantic map also suggests a strong connection between accounting and ethics topics, which means that accounting and audit studies on fraud often include ethical issues. The connection between accounting and internet fraud topics is, on the contrary, weak, indicating a gap in the literature than needs to be addressed.

Analysis of the semantic map also revealed the emergence of the topic of data driven fraud detection methods. This topic is expanding fast and appears in the works of both accounting and technology fraud researchers. Development of effective and reliable fraud detection methods seems to become a new hot topic in the fraud research. Data driven methods have the capacity to change the accounting and audit practices dramatically, and their growing use for fraud detection is definitely worth future studies.

We plan to continue our project and run semantic analysis on more specialized data sets. Our next step is semantic analysis of EBSCO hosted accounting academic journals. The semantic map presented in this paper shows most of the accounting literature as one giant node, "accounting and audit". Performing semantic analysis on accounting literature alone will allow for a more granular analysis of fraud related topics within this literature.

References

Albrecht, W. S., C. C. Albrecht, and C. O. Albrecht. 2006. *Fraud Examination*. New York, NY: Thomson South-Western.

Beasley, M. S. 1996. "An empirical analysis of the relation between the board of director composition and financial statement fraud," *Accounting review (71:4)*, pp. 443–465.

Blondel, V.B., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). "Fast unfolding of communities in large networks," *J. of Statistical Mechanics: Theory and Experiment, 1000.*

Brody, R. G., Melendy, S. R., and Perri, F. S. 2012. "Commentary from the American Accounting Association's 2011 Annual Meeting Panel on Emerging Issues in Fraud Research," *Accounting Horizons* (26:3), pp. 513–531.

Cooper, D. J., Dacin, T., and Palmer, D. 2013. "Fraud in accounting, organizations and society: Extending the boundaries of research," *Accounting, Organizations and Society* (38:6–7), pp. 440–457.

Darby, M. R., and Karni, E. 1973. "Free competition and the optimal amount of fraud," *The Journal of law and economics* (16:1), pp. 67–88.

Dorminey, J., Fleming, A. S., Kranacher, M.-J., and Riley Jr., R. A. 2012. "The Evolution of Fraud Theory," *Issues in Accounting Education* (27:2), pp. 555–579.

Fawcett, T., and Provost, F. 1997. "Adaptive fraud detection," *Data mining and knowledge discovery* (1:3), pp. 291–316.

Hogan, C. E., Rezaee, Z., Riley Jr, R. A., and Velury, U. K. 2008. "Financial statement fraud: Insights from the academic literature," *Auditing: A Journal of Practice & Theory* (27:2), pp. 231–252.

Kranacher, M. J., R. A. Riley Jr., and J. T. Wells. 2011. *Forensic Accounting and Fraud Examination.* New York, NY: John Wiley & Sons

Pérez-Corona, N., Hernández-Colín, D. Bustillo-Hernández, C., and Figueroa-Nazuno, J. (2012) "Model of natural semantic space for ontologies' construction," *International Journal of Combinatorial Optimization Problems & Informatics, (*3:2), 93-108.

Public Company Accounting Oversight Board (PCAOB). 2004. PCAOB Standing Advisory Group Meeting: Financial Fraud, September 8–9. Available at: *https://pcaobus.org/News/Events/Documents/09082004_SAGMeeting/Fraud.pdf*

Ramamoorti, S. 2008. "The psychology and sociology of fraud: Integrating the behavioral sciences component into fraud and forensic accounting curricula," *Issues in Accounting Education* (23:4), pp. 521–533.

Rice, R.E. and Danowski, J.A. (1993), "Is it really just like a fancy answering machine? Comparing semantic networks of different types of voice mail users", *The Journal of Business Communications,* (30:4), pp. 369-97.

Rust, Roland T. and Ming-Hui Huang (2014), "The service revolution and the transformation of marketing science," *Marketing Science, (*33:2), pp.206-221.

Salton, G., Wong, A., and Yang, C.S. (1975). "A vector space model for automatic indexing," *Communications of ACM, (*18:11), pp. 613-620.

Steyvers, M. and Tenenbaum, J. (2005). "The large-scale structure of semantic networks: statistical analyses and a model of semantic growth," *Cognitive Science, (*29), pp.41-78.

Trompeter, G. M., Carpenter, T. D., Desai, N., Jones, K. L., and Riley Jr, R. A. 2013. "A synthesis of fraud-related research," *Auditing: A Journal of Practice & Theory* (32:sp1), pp. 287–321.

Trompeter, G. M., Carpenter, T. D., Jones, K. L., and Riley Jr., R. A. 2014. "Insights for Research and Practice: What We Learn about Fraud from Other Disciplines," *Accounting Horizons* (28:4), pp. 769–804.