THE JOURNAL OF
DIGITAL FORENSICS,
SECURITY AND LAW

Journal of Digital Forensics,
Security and Law

Volume 12 | Number 2

Article 8

6-30-2017

Towards a More Representative Definition of Cyber Security

Daniel Schatz
University of East London, u0829943@uel.ac.uk

Rabih Bashroush
University of East London, r.bashroush@uel.ac.uk

Julie Wall
University of East London, j.wall@uel.ac.uk

Recommended Citation

Schatz, Daniel; Bashroush, Rabih; and Wall, Julie (2017) "Towards a More Representative Definition of Cyber Security," *Journal of Digital Forensics, Security and Law*: Vol. 12 : No. 2 , Article 8.
DOI: https://doi.org/10.15394/jdfsl.2017.1476
Available at: https://commons.erau.edu/jdfsl/vol12/iss2/8

TOWARDS A MORE REPRESENTATIVE DEFINITION OF CYBER SECURITY

Daniel Schatz
University of East London
Architecture Computing and Engineering
u0829943@uel.ac.uk

Rabih Bashroush
University of East London
Architecture Computing and Engineering
r.bashroush@uel.ac.uk

Julie Wall
University of East London
Architecture Computing and Engineering
j.wall@uel.ac.uk

ABSTRACT

In recent years, 'Cyber Security' has emerged as a widely-used term with increased adoption by practitioners and politicians alike. However, as with many fashionable jargon, there seems to be very little understanding of what the term really entails. Although this is may not be an issue when the term is used in an informal context, it can potentially cause considerable problems in context of organizational strategy, business objectives, or international agreements. In this work, we study the existing literature to identify the main definitions provided for the term 'Cyber Security' by authoritative sources. We then conduct various lexical and semantic analysis techniques in an attempt to better understand the scope and context of these definitions, along with their relevance. Finally, based on the analysis conducted, we propose a new improved definition that we then demonstrate to be a more representative definition using the same lexical and semantic analysis techniques.

Keywords: cyber security; information security; national cyber policy; systematic review

1. INTRODUCTION

The terminology used to discuss security aspects of digital devices and information changed considerably in recent years. At the beginning of the century, terms regularly used in this context would be "Computer Security," "IT Security," or "Information Security." Whilst these terms have nuanced differences understood by professionals working in this space, they were tangible enough to be meaningful to the wider populace. General conversations could be had and plans could be made based on a common understanding of what these terms imply. However, towards the end of the first decade, new terminology started to become increasingly popular with the use of the term "Cyber Security." It had been in use during previous years but its popularity gained considerably when U.S. President Barack Obama in 2009 proclaimed "I call upon the people of the United States to recognize the importance of cybersecurity and to observe this month with appropriate activities, events, and trainings to enhance our national security and resilience" (The White House, 2009). The immediate impact of this press release on terminology can be illustrated with the help of Google's search trends which shows a noticeable spike in this period (Figure 1). The trend lines on the chart show total searches for a term relative to the total number of searches done on Google over time. We can see a steady decline of the search terms "Computer Security" and "Information Security" with variants of "Cyber Security" converging and surpassing them. This finding is only indicative but as seen in previous research (Choi & Varian, 2012), search engine based information is useful and of value to identify trends.

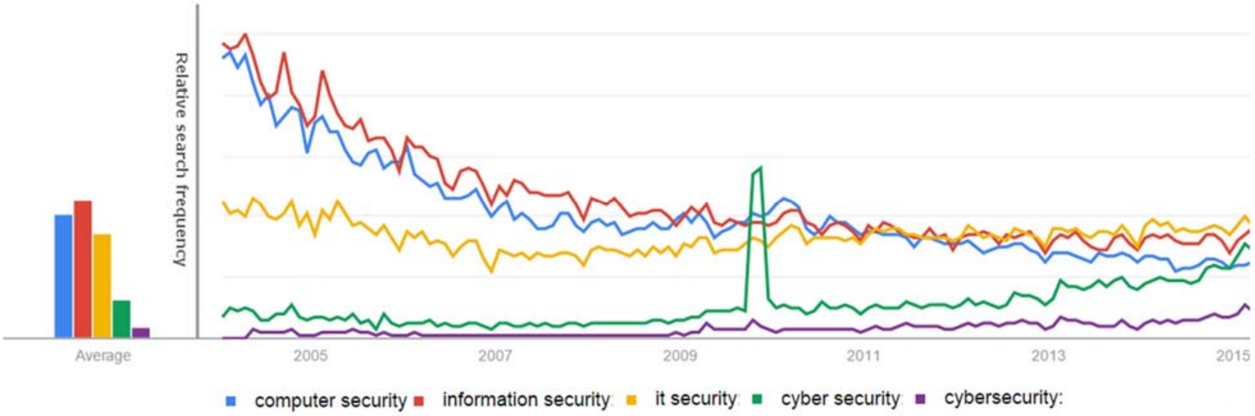


Figure 1. Google search trends for security 2004 – 2015

This development in use of terminology is causing some issues as the term "Cyber Security" lacks the defining clarity of, for example, "Computer Security." This can lead to confusion and misunderstanding if parties have different assumptions of what the term represents. Quoting Sowell (2014) on the importance of clarity;

> "What may seem like small steps in logic, after the fact, can be a long, time-consuming process of trial and error groping, while creating and refining concepts and definitions to express ideas in clear and unmistakable terms which allow substantive issues to be debated in terms that opposing parties can agree on, so that they can at least disagree on substance, rather than be frustrated by semantics."

Whilst it is unlikely to be a problem in private conversations between interested citizens it becomes, at the very least, a nuisance at an organizational level and is a widely recognized issue among professionals in the field. These problems amplify if ambiguity continues in courts of justice, national cyber security strategies or international treaties. Eig (2011) discusses such issues in context of statutory interpretation in greater detail.

An additional, although less impacting, issue is the inconsistent use of syntax for cyber security. Across the literature both versions, cybersecurity and cyber security, are used. Observing the search trends as illustrated we see that both terms are upward trending; however, the disjoined version (cyber security) shows prevalence in absolute numbers which is the spelling that shall be used going forward unless referring to primary source material.

Recognizing the lack of a consistent meaning of the term cyber security as a considerable issue (Baylon, 2014; Congressional Research Service, 2014; Creasey, 2013; Internet Society, 2012), we are first reviewing the current definition landscape across professional, academic and governmental literature with a goal to identify the most prevalent definitions, key components in definitions, and take a view on contentious points between proposed definitions where such exist. As second and third steps, we will identify the best match definition and contribute a new improved one.

The remainder of the paper is structured as follows; in the next section, we will take a look at existing research in this field and discuss challenges of the current definition landscape. Section 3 describes the approach and methodology followed for our systematic literature review on the topic. We continue to analyze the definition set from a semantic perspective in sections 4 and 5 with a proposal for an improved definition outlined in section 6. In sections 7 and 8, we review limitations of our approach and provide conclusive thoughts.

2. LITERATURE REVIEW

The lack of a uniformly accepted definition of cyber security as described in the previous section has been recognized across professional (Barzilay, 2013; Stubley, 2013; Walls, Perkins, & Weiss, 2013), governmental (Falessi, Gavrila, Klejnstrup Ritter, & Moulinos, 2012; Government of Montenegro, 2013; Wamala, 2011) and academic (Baylon, 2014; Giles & Hagestad, 2013) work.

2.1 Industry definitions

Walls et al. (2013) approach the topic from the perspective of a professional services provider (Gartner Inc.) and is thus focusing on tangible guidance for strategic decision makers. A key challenge highlighted is the ambiguity introduced by the thoughtless use of the term 'cyber security,' where nuanced definitions (Information Security or IT Security) are more appropriate and descriptive. They suggest that the term cyber security is only used in context of security practices related to the combination of offensive and defensive actions involving or relying upon information technology and/or operational technology environments and systems. The authors state that it marks a superset of security practices such as information security, IT security and other related practices. Stubley (2013) takes a different view to this and simplifies cyber security to information security based on a short analysis of the 'cyber' component which he defines to describe the use of information technology and computers. Barzilay (2013) again takes a different view and argues that cyber security must be defined through cyber risk which leads to his conclusion that cyber security is a sub discipline of information security which is in contrast to Walls et al. (2013). In official guidance, ISACA (2014) takes yet another position stating that cyber security is emerging within the fields of information security and traditional security. Enterprises should distinguish between standard (lower-level) information security and cyber security; the difference is in the scope, motive, opportunity and method of the attack.

2.2 Government and nation state definitions

In their analysis of national cyber security strategies of European Union member states, Falessi et al. (2012) provide terminology guidance in the annex explaining that there is no universally accepted nor straightforward definition of 'cyber security.' They write that some people regard cyber security as overlapping with information security but no definitive conclusion is provided. This view is shared by Wamala (2011) claiming that cyber security is a branch of information security. The paper highlights the risk of uncertain terminology and aims to provide clarification on the relative positions of cyber security and information security. It draws a link between cyber security and the global characteristic of the internet, as such distinguishing it from information security which, according to the author, rarely traverses jurisdictions. Wamala goes further in this definition claiming that cyber security focuses more on integrity and availability whereas information security is mainly concerned with confidentiality. He concludes that cyber security is information security with jurisdictional uncertainty and attribution issues. The Government of Montenegro (2013) agrees with the notion of a lack of clear definitions in this area and dedicates a full section in its cyber security strategy to this topic. Whilst the paper states that it presents definitions which are compliant with the basic meanings as understood in EU countries, it unfortunately does not actually provide a conclusion on the term cyber security but rather quotes various definitions from other sources. Baylon (2014) discusses the topic from a multinational cooperation perspective highlighting that the lack of or insufficiently agreed on definition of key terminology in the cyber and space security domains poses a major challenge to international treaties and arms control agreements. In particular, the considerably different interpretation of cyber security between western countries and both Russia and China causes complications in this context. Baylon states that the term 'cyber security' as such does not exist in Russian legislation or official doctrines. Instead, the concept of information security is prevalent. However, in this context, "information" represents a meaning extending outside the digital space which widens conversations into the information space in general. The author categorizes this into the Eastern approach, looking at cyber security emphasizing 'social cohesion,' and the Western approach, perceiving cyber security through a 'national security prism.' Godwin III, Kulpin, Rauscher, and Yaschenko (2014) concur with this challenge and provide bi-nationally (USA/Russia) agreed terminology for key phrases pertaining to the cyber space. Amongst these, the term 'cyber security' is defined as well; notably with a considerably different interpretation than found in official national cyber security strategies of most western countries. Giles and Hagestad (2013) extend on this by contrasting key terms and principles in this space as understood in each of their focus countries (USA, China, Russia). They find that there is a notably different understanding and approach between these countries. They conclude that in absence of a mutually agreed terminology, any potential for finding a real commonality of views on the nature and governance of cyberspace remains distant.

2.3 Academic definitions

Academic research has not been oblivious to the obvious challenges in this developing problem space, of course. Luiijf, Besseling, and de Graaf (2013) conducted an exhaustive study of national cyber security strategies (NCSS) for 19 countries which also discusses differences in terminology in some detail. They find that only eight nations define the term 'cyber security' in their NCSS, whereas six nations do not provide any such definition. The authors note that for the ten NCSS which have the term cyber security defined either through implication, description or definition, the understanding of what it means varies greatly. This view is shared by Craigen, Diakun-Thibault, and Purse (2014) who looked at a wider range of sources attempting to define the term. They find that the term is used broadly and its definitions are highly variable, context-bound, often subjective, and, at times, uninformative. Based on a shortlist of nine definitions, the authors work towards a unified definition identifying five dominant themes of cyber security. Through consensus in a multidisciplinary group, the authors arrive at an additional definition for cyber security. Many of the definitions mentioned in this section will be the focus of the remainder of this paper.

3. SYSTEMATIC REVIEW APPROACH

To better understand the variety of relevant definitions in use for cyber security, we followed a semi systematic literature review approach (Mäntylä, Adams, Khomh, Engström, & Petersen, 2014) as further described below. Following the collection of definitions, we applied text analysis methods on the resulting dataset focusing on semantic similarity analysis with the goal to identify harmonizing definitions. This approach resulted in a ranking of definition similarity across the dataset from a text analytics perspective; i.e. we established which definitions represent most accurately the definition of 'cyber security' across the whole dataset. Based on further analysis of the highest scoring definitions, we created a new definition comprising the key terms identified. The new definition was then compared against the original dataset to verify its best match status across the whole dataset.

3.1 Research question

We started by defining our research questions at a high level.

Table 1
Research questions

RQ 1	What definitions are currently used for 'cyber security' by authoritative sources?
	The intention is to understand how cyber security is currently defined by sources of authority (academic, professional, government)
RQ 2	Are there differences in the definitions?
	The intention is to understand whether the definitions are similar or considerably different
RQ 3	Is there a best match definition of cyber security
	The assumption is that there are various definitions proposed so we're trying to identify the best match definition across the dataset
RQ 4	Are we able to contribute a new best match definition of cyber security
	Based on a text analysis approach, are we able to provide a new best match definition?

In order to answer our research questions, we first needed to identify the relevant definitions. For this, we applied a set of inclusion and exclusion criteria to our literature search as follows.

Inclusion criteria:

- **IC1**: Sources with clear intention of providing an explicit definition of cyber security
- **IC2**: Sources available in English or translation readily available
- ***Exclusion criteria:***
- **EC1**: Sources which provide no clear or only implicit definitions of cyber security
- **EC2**: Sources that lack rigor (peer review) or authority (governmental or professional bodies) for defining cyber security

These criteria have been applied throughout the search process, in particular EC2 (Cornell University, 2016). In the first instance, Thomson Reuters' Web of Science database was used to identify relevant academic sources. The search scope covered a time span of 'All years' with a search construct of TOPIC: (("cyber security" OR Cybersecurity) NEAR definition). This produced limited results of merely 13 hits of which only one source met our criteria. Modifying the search query to include variations of the term 'definition' (meaning, interpretation) did not produce any additional relevant results. Our search efforts in other databases such as Science Direct (25 results) were met with similar challenges. To capture a wider range of sources we extended our search efforts to the general purpose search engine Google.com, limiting search parameters as follows ([cybersecurity AROUND(3) definition] OR ["cyber security" AROUND(3) definition]). Manual review of the top search results returned by Google was then conducted to capture the most relevant sources. Based on the sources identified, further backward and forward reference crawling was conducted (using Google Scholar) to capture additional material relevant to our research question. In addition, source lists provided by ENISAi and NATO ii were reviewed manually. Our literature review identified 28 sources which met our inclusion and exclusion criteria as shown in Appendix A in no particular order. Out of the 28 identified sources, one definition source is considered academic, five industries contributed and 22 definitions were by government or government aligned bodies. As expected there is considerable overlap in term use between definitions of which some include parts of definitions stated by another source (e.g. #3 and #18). The definition text was extracted from the source material in the context it was written.

4. BASIC DEFINITION ANALYSIS

To get a better understanding of the dataset, an initial exploratory text analysis (Hearst, 1999) was conducted to try and discover information inherent to the definitions. We started by applying basic information extraction procedures (Weiss, Indurkhya, Zhang, & Damerau, 2004) utilizing the Text Mining framework *tm_ map* (Meyer, Hornik, & Feinerer, 2008) for the software environment for statistical computing "R."

Before the definition data were loaded into R, minimal manual normalization was applied to standardize character encoding and remove unnecessary line breaks. The definition corpus was then prepared with common pre-processing functions as provided by tm_map to convert content to lower case, strip whitespaces, remove punctuation and remove stop words (English). In addition, stemming was applied (Porter, 1997) to reduce the number of distinct word types in the text corpus and to increase the frequency of occurrence of some individual types (Weiss et al., 2004).

With the corpus prepared, we created a simple document-term matrix (Salton, 1963) that allowed us to gain basic insights on how our sources define 'cyber security.' As illustrated in Figure 2, the root form of ‘security,’ ‘cyber security,’ ‘cyber,’ and ‘cyberspace’ is prevalent in the corpus which was expected. However, we also get an indication of related words fundamental to the definition pool.

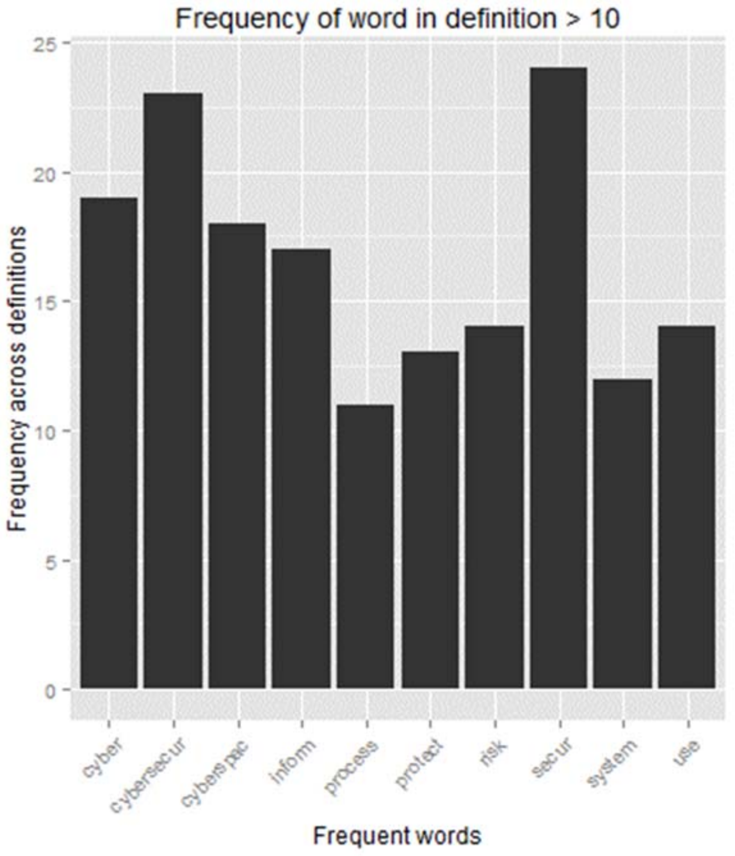
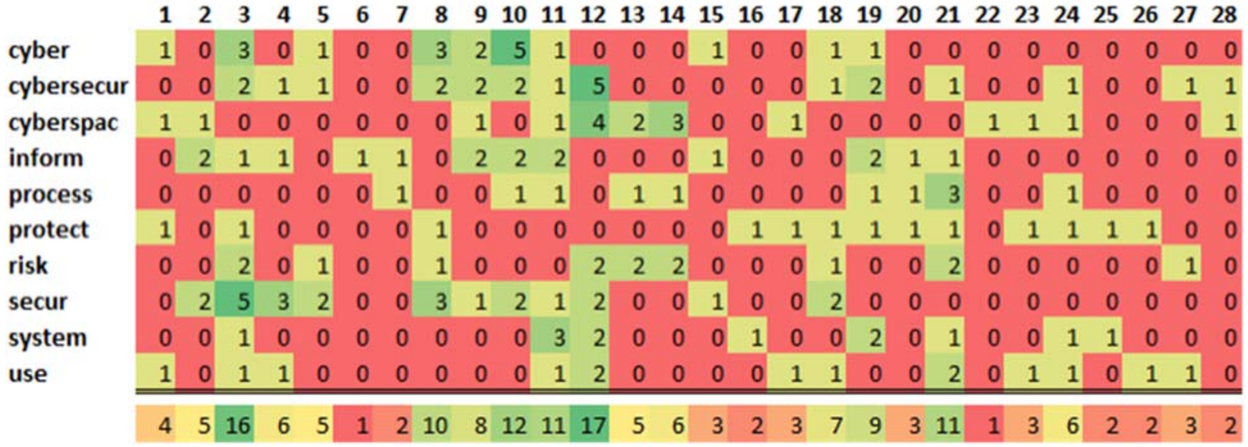


Figure 2. Word frequency in definition corpus for top terms

The basic term frequency analysis provided an intuition on term priority across the definition dataset and an indication of the importance (by way of word count) of certain words in the set, most notably ‘risk,’ ‘protect,’ ‘use,’ ‘process,’ and ‘system.’ With this information, we conducted an analysis on the definition sets.

Lexical Overlap analysis, the process of identifying how many words texts have in common, is one of the simplest methods to assess the similarity between texts (Rus, 2014). We used this to conduct a basic lexical token review on our definition set with just the most frequent ten unigrams in their stemmed form as shown in Figure 3. The heat map shows the ten most frequent terms across all 28 definitions in the dataset with an individual and total term count per document. Based on this simple analysis, we glean that some definitions incorporate a wider spectrum of key terms than others (e.g. #3 or #11), and may provide a better representation of what the entire definition pool defines as ‘cyber security.’ We also note that such simple analysis is skewed by term repetition e.g. as observed for definition #12 where 'cybersecur' and 'cyberspac' are used frequently.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
cyber	1	0	3	0	1	0	0	3	2	5	1	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0
cybersecur	0	0	2	1	1	0	0	2	2	2	1	5	0	0	0	0	0	1	2	0	1	0	0	1	0	0	1	1
cyberspac	1	1	0	0	0	0	0	0	1	0	1	4	2	3	0	0	1	0	0	0	0	1	1	1	0	0	0	1
inform	0	2	1	1	0	1	1	0	2	2	2	0	0	0	1	0	0	0	2	1	1	0	0	0	0	0	0	0
process	0	0	0	0	0	0	1	0	0	1	1	0	1	1	0	0	0	0	1	1	3	0	0	1	0	0	0	0
protect	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	1	1	1	1	1	0	1	1	1	1	0	0
risk	0	0	2	0	1	0	0	1	0	0	0	2	2	2	0	0	0	1	0	0	2	0	0	0	0	0	1	0
secur	0	2	5	3	2	0	0	3	1	2	1	2	0	0	1	0	0	2	0	0	0	0	0	0	0	0	0	0
system	0	0	1	0	0	0	0	0	0	0	3	2	0	0	0	1	0	0	2	0	1	0	0	1	1	0	0	0
use	1	0	1	1	0	0	0	0	0	0	1	2	0	0	0	0	1	1	0	0	2	0	1	1	0	1	1	0
	4	5	16	6	5	1	2	10	8	12	11	17	5	6	3	2	3	7	9	3	11	1	3	6	2	2	3	2

Figure 3. Heat map of the most frequent word stem analysis across 28 definition sources

Continuing with our basic analysis, we created a correlation matrix for a sparse document-term matrix (sparsity at 0.85) to gain additional information on strongly correlated terms across the definition set. Figure 4 confirms our assumption that frequent terms such as 'cybersecur,' 'cyberspac,' or 'secur' are not highly correlated with other terms in this context; however, the correlation matrix shows that we have some correlated terms that are worth further exploration.

We found high correlation between the terms of the CIA triad (confidentiality, integrity, availability) which makes intuitive sense as they tend to be used together when writing about topics like cyber security. We further see noteworthy correlation of 'inform' and 'integr' along with the CIA triad which we'll see confirmed in a later section of this paper. We also note correlation between 'secur,' 'asset,' and 'environ' which points towards a general agreement that those terms standing together are important to a harmonized definition of cyber security. This basic approach shows further interesting positive and negative correlations (e.g. 'include' and 'infrastructur' or 'realibl' and 'protect') that helps to better understand the definition space. But we still lack a way to identify what the most representative definition of 'cyber security' is. Following the maxim "a person without data is just another person with an opinion"[iii], we designed an approach that would allow us to identify the most representative definition within our pool. The assumption is that our dataset includes the majority of authoritative definitions for 'cyber security' and as such covers all relevant aspects of the concept as proposed by the sources (Ryan & Bernard, 2003). This means we can identify the definition encompassing the majority of relevant components through lexical and semantic similarity analysis; that is the definition which is most alike to every other definition in the dataset. We made use of a wide range of advanced similarity measures as described in the next section to achieve this.

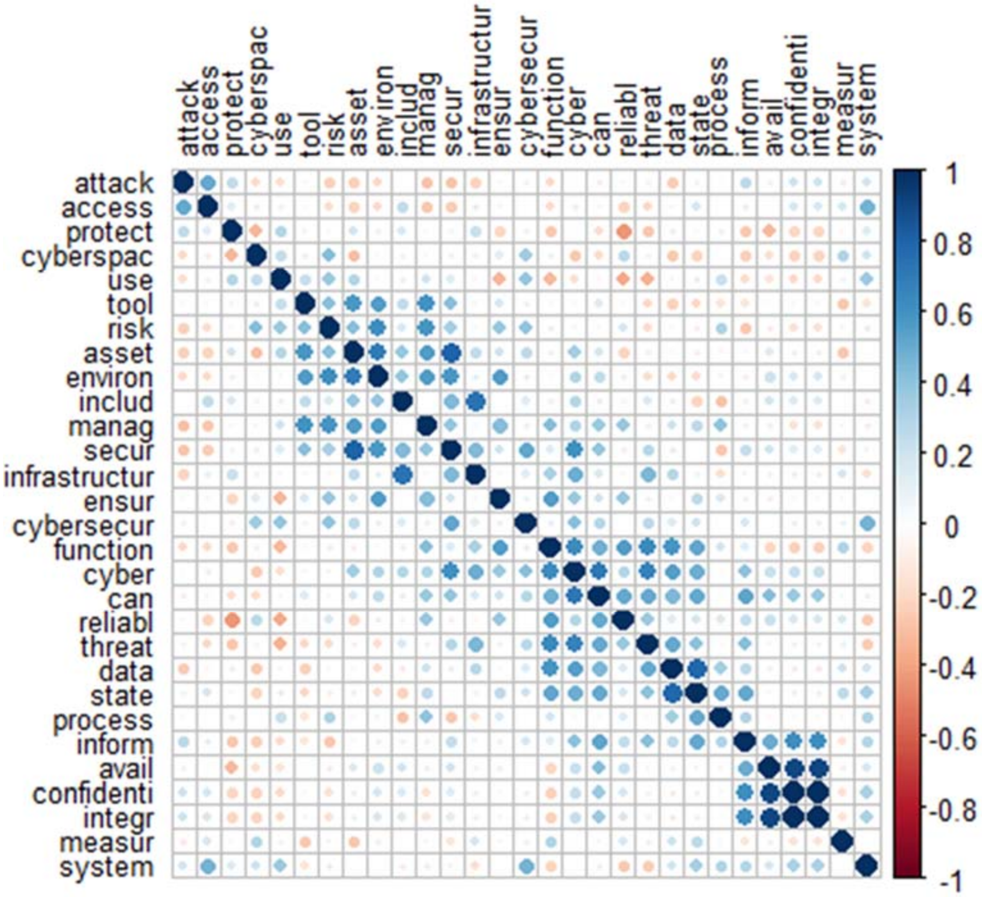


Figure 4. Sparse DTM correlation matrix

5. DEFINITION SIMILARITY ANALYSIS

Semantic similarity is a well-established area of research with a wide range of practical applications (Androutsopoulos & Malakasiotis, 2010; Couto, Silva, & Coutinho, 2007; Graesser, Olney, Haynes, & Chipman, 2005; Yuhua, Bandar, & McLean, 2003). For the purpose of this research, we investigated current work on short text and sentence based similarity measures. We initially planned to use the best method for sentence based similarity measures as proposed by subject matter experts on this topic, but found that this is a developing area with various competing methods proposed. Instead of picking one specific method to calculate similarity, we decided to calculate similarity with a variety of methods to balance advantages and disadvantages of individual methods. The result is an average similarity score as described in this section. We found the SEMILAR toolkit (Rus, Lintean, Banjade, Niraula, & Stefanescu, 2013) to be ideal for this as it vastly simplified the task of calculating similarity using multiple algorithms and options. The authors describe the toolkit as “a one-stop-shop for investigating, annotating, and authoring methods for the semantic similarity of texts of any level of granularity.” We used the toolkit to conduct both the pre-processing phase and the similarity computing phase for our dataset.

As with our basic analysis we conducted common pre-processing tasks on our dataset but with some notable differences. Again, the first step is tokenization of the text to obtain the ordered set of lexical tokens. Based on our configuration, SEMILAR calculates the initial lexical form of the token, lemma form of the word, part-of-speech (POS), weighted specificity of the word, semantic representation (WordNet (Miller, 1995) or LSA (Martin & Berry, 2007)) and a list of syntactic dependencies with the other words in the same sentence (Lintean, 2011). To capture as much context as possible, we chose Stanford CoreNLP (De Marneffe, MacCartney, & Manning, 2006) as the configuration option for tokenization, POS tagging, lemmatizer as well as syntactic parsing. Figure 5 provides a visual example of how this task processed one of the definitions in the set. The effect of lemmatization (as compared to stemming) and part of speech tagging is apparent. The function identified sentence tokens and categorized them accurately for further processing. In the sample chosen we see that the tagger associated words with their respective part of speech (“The” /Determiner, “ability” /Noun singular, “protect” /verb base, “or” /coordinating conjunction, etc.).

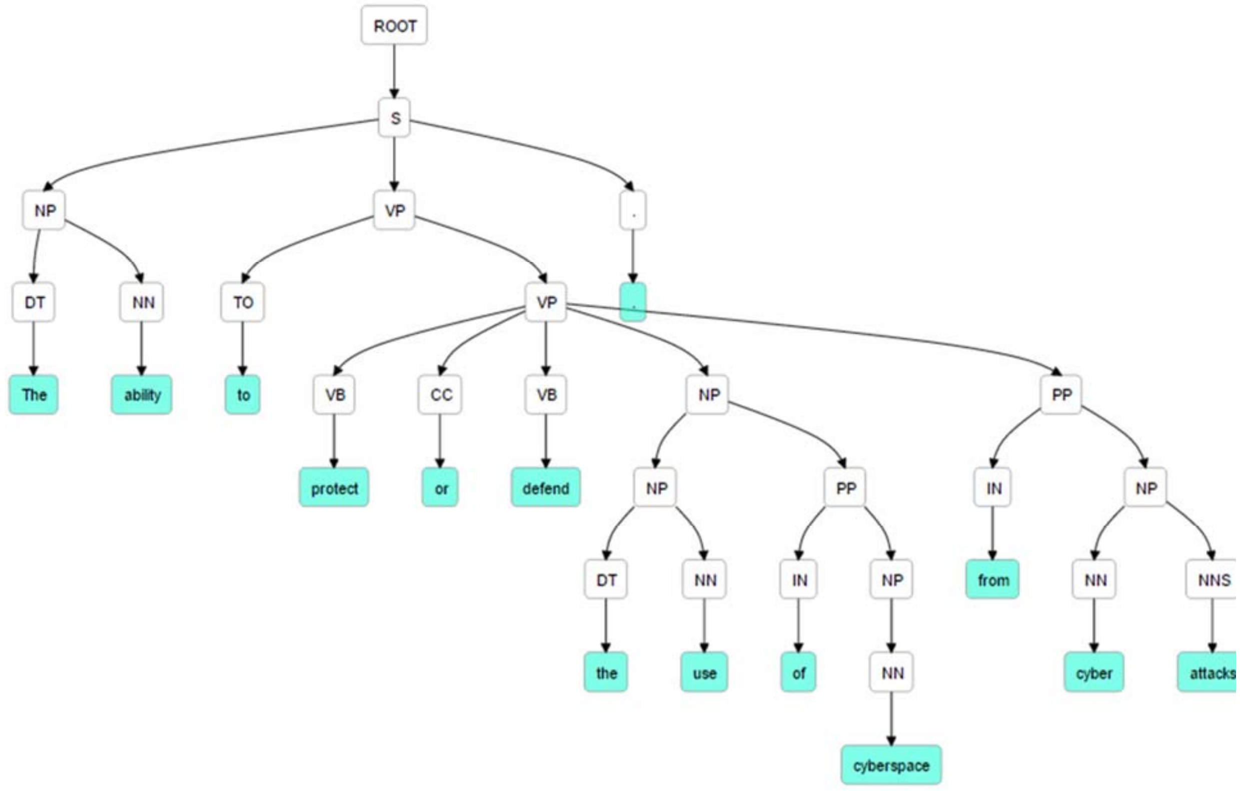


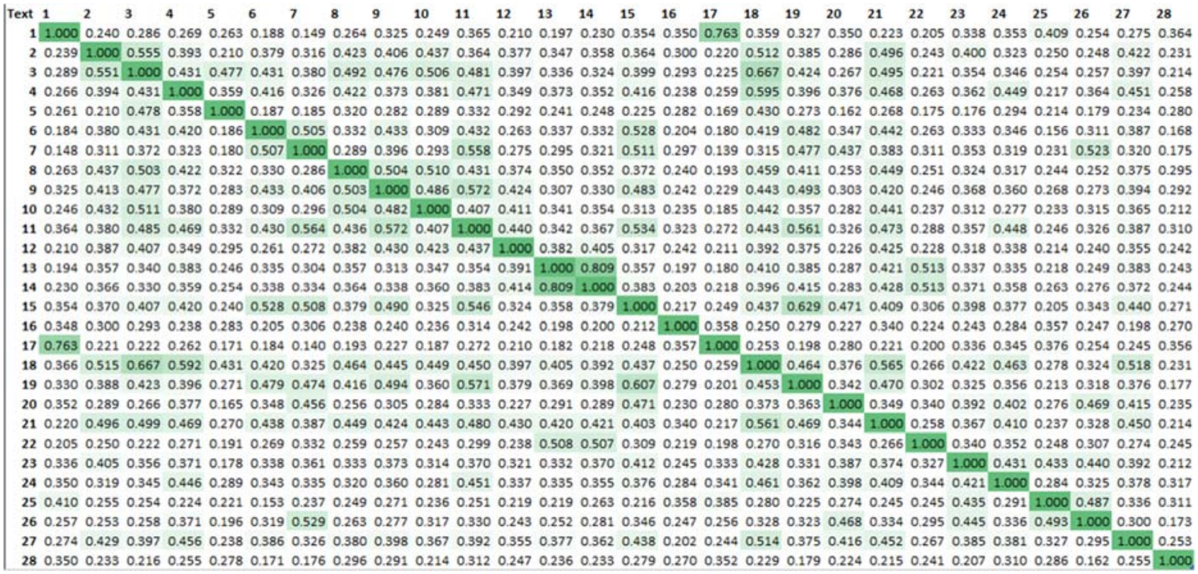
Figure 5. Sample pre-processing with StanfordNLP

With the definition set prepared this way, we calculated similarity between all definitions using nine methods resulting in 7056 similarity scores. The selection of the nine methods and their configuration options we used to calculate the similarity scores was based on recommendations and insights in relevant literature (Corley & Mihalcea, 2005; Gomaa & Fahmy, 2013; Lee, 2011; Lintean, 2011; Nakov, Popova, & Mateev, 2001; Rus, 2014; Rus & Lintean, 2012; Yuhua, McLean, Bandar, O'Shea, & Crockett, 2006). The methods chosen are categorized and listed within the SEMILAR toolkit as lexical methods (five), Corley and Mihalcea (2005) (three) as well as plain LSA vector similarity.

For lexical similarity methods, we did not remove stop words, non-function words or punctuation adopting findings by Lintean (2011, p. 60) and Yuhua et al. (2006) showing the importance of these tokens for similarity calculations due to their structural information value. We did, however, convert all tokens to lower case. For lexical matching we selected optimal pairing (Rus et al., 2012) without enforcing part of speech matching. Token weights are based on entropy (Martin & Berry, 2007) rather than inverted document frequency (IDF) (Sparck Jones, 1972) following guidance by Lintean (2011), finding that entropy-based weighting leads to better results than IDF-based weighting. With this configuration set as baseline, we selected five token similarity metric methods; Jiang and Conrath (Jiang & Conrath, 1997), Leacock and Chodorow (Leacock, Miller, & Chodorow, 1998), Lin (Lin, 1998), LSA as well as Wu and Palmer (Wu & Palmer, 1994). For similarity calculations based on the 'Class of Method' (Corley & Mihalcea), we again chose Jiang and Conrath, Leacock and Chodorow and Lin, each with bidirectional scoring type and Touchstone Applied Science Associates corpus (TASA) derived IDF as model. Finally, for plain LSA similarity scoring, we selected a frequency-based local weight, as well as an entropy-based global weight. Further rationale and detail on each of the methods is beyond the scope of this paper and can be found in the references listed in this section.

With the scores calculated, we transformed them into a matrix format as pictured in

Figure *6*. This allowed us to calculate final averages for each definition.



Text	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1	1.000	0.240	0.286	0.269	0.263	0.188	0.149	0.264	0.325	0.249	0.365	0.210	0.197	0.230	0.354	0.350	0.763	0.359	0.327	0.350	0.223	0.205	0.338	0.353	0.409	0.254	0.275	0.364
2	0.239	1.000	0.555	0.393	0.210	0.379	0.316	0.423	0.406	0.437	0.364	0.377	0.347	0.358	0.364	0.300	0.220	0.512	0.385	0.286	0.496	0.243	0.400	0.323	0.250	0.248	0.422	0.231
3	0.289	0.551	1.000	0.431	0.477	0.431	0.380	0.492	0.476	0.506	0.481	0.397	0.336	0.324	0.399	0.293	0.225	0.667	0.424	0.267	0.495	0.221	0.354	0.346	0.254	0.257	0.397	0.214
4	0.266	0.394	0.431	1.000	0.359	0.416	0.326	0.422	0.373	0.381	0.471	0.349	0.373	0.352	0.416	0.238	0.259	0.595	0.396	0.376	0.468	0.263	0.362	0.449	0.217	0.364	0.451	0.258
5	0.261	0.210	0.478	0.358	1.000	0.187	0.185	0.320	0.282	0.289	0.332	0.292	0.241	0.248	0.225	0.282	0.169	0.430	0.273	0.162	0.268	0.175	0.176	0.294	0.214	0.179	0.234	0.280
6	0.184	0.380	0.431	0.420	0.186	1.000	0.505	0.332	0.433	0.309	0.432	0.263	0.337	0.332	0.528	0.204	0.180	0.419	0.482	0.347	0.442	0.263	0.333	0.346	0.156	0.311	0.387	0.168
7	0.148	0.311	0.372	0.323	0.180	0.507	1.000	0.289	0.396	0.293	0.558	0.275	0.295	0.321	0.511	0.297	0.139	0.315	0.477	0.437	0.383	0.311	0.353	0.319	0.231	0.523	0.320	0.175
8	0.263	0.437	0.503	0.422	0.322	0.330	0.286	1.000	0.504	0.510	0.431	0.374	0.350	0.352	0.372	0.240	0.193	0.459	0.411	0.253	0.449	0.251	0.324	0.317	0.244	0.252	0.375	0.295
9	0.325	0.413	0.477	0.372	0.283	0.433	0.406	0.503	1.000	0.486	0.572	0.424	0.307	0.330	0.483	0.242	0.229	0.443	0.493	0.303	0.420	0.246	0.368	0.360	0.268	0.273	0.394	0.292
10	0.246	0.432	0.511	0.380	0.289	0.309	0.296	0.504	0.482	1.000	0.407	0.411	0.341	0.354	0.313	0.235	0.185	0.442	0.357	0.282	0.441	0.237	0.312	0.277	0.233	0.315	0.365	0.212
11	0.364	0.380	0.485	0.469	0.332	0.430	0.564	0.436	0.572	0.407	1.000	0.440	0.342	0.367	0.534	0.323	0.272	0.443	0.561	0.326	0.473	0.288	0.357	0.448	0.246	0.326	0.387	0.310
12	0.210	0.387	0.407	0.349	0.295	0.261	0.272	0.382	0.430	0.423	0.437	1.000	0.382	0.405	0.317	0.242	0.211	0.392	0.375	0.226	0.425	0.228	0.318	0.338	0.214	0.240	0.355	0.242
13	0.194	0.357	0.340	0.383	0.246	0.335	0.304	0.357	0.313	0.347	0.354	0.391	1.000	0.809	0.357	0.197	0.180	0.410	0.385	0.287	0.421	0.513	0.337	0.335	0.218	0.249	0.383	0.243
14	0.230	0.366	0.330	0.359	0.254	0.338	0.334	0.364	0.338	0.360	0.383	0.414	0.809	1.000	0.383	0.203	0.218	0.396	0.415	0.283	0.428	0.513	0.371	0.358	0.263	0.276	0.372	0.244
15	0.354	0.370	0.407	0.420	0.240	0.528	0.508	0.379	0.490	0.325	0.546	0.324	0.358	0.379	1.000	0.217	0.249	0.437	0.629	0.471	0.409	0.306	0.398	0.377	0.205	0.343	0.440	0.271
16	0.348	0.300	0.293	0.238	0.283	0.205	0.306	0.238	0.240	0.236	0.314	0.242	0.198	0.200	0.212	1.000	0.358	0.250	0.279	0.227	0.340	0.224	0.243	0.284	0.357	0.247	0.198	0.270
17	0.763	0.221	0.222	0.262	0.171	0.184	0.140	0.193	0.227	0.187	0.272	0.210	0.182	0.218	0.248	0.357	1.000	0.253	0.198	0.280	0.221	0.200	0.336	0.345	0.376	0.254	0.245	0.356
18	0.366	0.515	0.667	0.592	0.431	0.420	0.325	0.464	0.445	0.449	0.450	0.397	0.405	0.392	0.437	0.250	0.259	1.000	0.464	0.376	0.565	0.266	0.422	0.463	0.278	0.324	0.518	0.231
19	0.330	0.388	0.423	0.396	0.271	0.479	0.474	0.416	0.494	0.360	0.571	0.379	0.369	0.398	0.607	0.279	0.201	0.453	1.000	0.342	0.470	0.302	0.325	0.356	0.213	0.318	0.376	0.177
20	0.352	0.289	0.266	0.377	0.165	0.348	0.456	0.256	0.305	0.284	0.333	0.227	0.291	0.289	0.471	0.230	0.280	0.373	0.363	1.000	0.349	0.340	0.392	0.402	0.276	0.469	0.415	0.235
21	0.220	0.496	0.499	0.469	0.270	0.438	0.387	0.449	0.424	0.443	0.480	0.430	0.420	0.421	0.403	0.340	0.217	0.561	0.469	0.344	1.000	0.258	0.367	0.410	0.237	0.328	0.450	0.214
22	0.205	0.250	0.222	0.271	0.191	0.269	0.332	0.259	0.257	0.243	0.299	0.238	0.508	0.507	0.309	0.219	0.198	0.270	0.316	0.343	0.266	1.000	0.340	0.352	0.248	0.307	0.274	0.245
23	0.336	0.405	0.356	0.371	0.178	0.338	0.361	0.333	0.373	0.314	0.370	0.321	0.332	0.370	0.412	0.245	0.333	0.428	0.331	0.387	0.374	0.327	1.000	0.431	0.433	0.440	0.392	0.212
24	0.350	0.319	0.345	0.446	0.289	0.343	0.335	0.320	0.360	0.281	0.451	0.337	0.335	0.355	0.376	0.284	0.341	0.461	0.362	0.398	0.409	0.344	0.421	1.000	0.284	0.325	0.378	0.317
25	0.410	0.255	0.254	0.224	0.221	0.153	0.237	0.249	0.271	0.236	0.251	0.219	0.219	0.263	0.216	0.358	0.385	0.280	0.225	0.274	0.245	0.245	0.435	0.291	1.000	0.487	0.336	0.311
26	0.257	0.253	0.258	0.371	0.196	0.319	0.529	0.263	0.277	0.317	0.330	0.243	0.252	0.281	0.346	0.247	0.256	0.328	0.323	0.468	0.334	0.295	0.445	0.336	0.493	1.000	0.300	0.173
27	0.274	0.429	0.397	0.456	0.238	0.386	0.326	0.380	0.398	0.367	0.392	0.355	0.377	0.362	0.438	0.202	0.244	0.514	0.375	0.416	0.452	0.267	0.385	0.381	0.327	0.295	1.000	0.253
28	0.350	0.233	0.216	0.255	0.278	0.171	0.176	0.296	0.291	0.214	0.312	0.247	0.236	0.233	0.279	0.270	0.352	0.229	0.179	0.224	0.215	0.241	0.207	0.310	0.286	0.162	0.255	1.000

Figure 6. Cyber security definitions average similarity score matrix

As the similarity score is asymmetric (Lintean, 2011) for some of the methods (Text A → Text B ≠ Text B → Text A) illustrated by the different values in the upper triangle compared to the lower triangle, we calculated all row (r) and column (c) means. The combined mean provided the overall similarity score per definition. Figure 7 shows how each definition measures up in similarity against all other definitions.

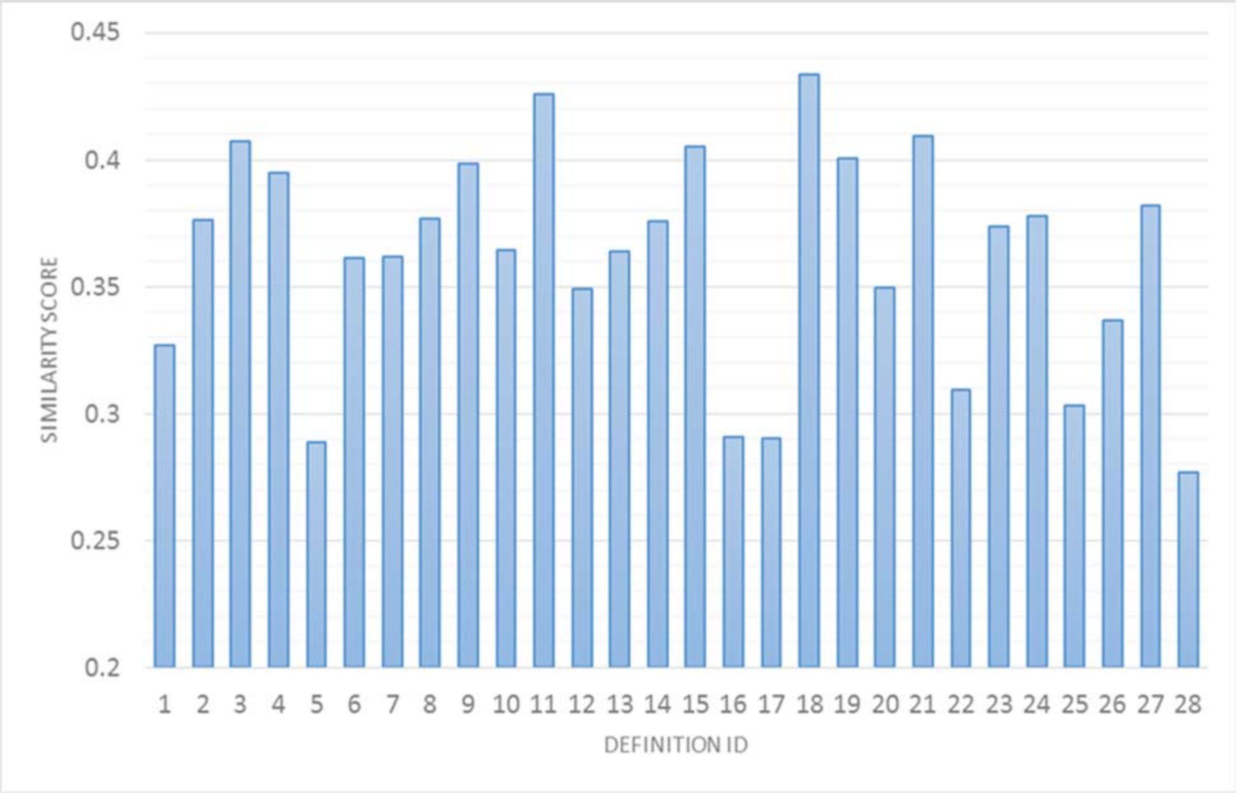


Figure 7. Average similarity score per definition in dataset

With this information, we produced a ranked order of the most representative definitions in the dataset. Table 2 shows an excerpt of the final list with the 5 most representative definitions of the definition pool ranked by similarity score across all nine methods and definitions.

Per our semantic similarity approach, the most representative definition in our dataset of authoritative definitions is part of the South Africa NCSS;

> "Cybersecurity is the collection of tools, policies, security concepts, security safeguards, guidelines, risk management approaches, actions, training, best practices, assurance and technologies that can be used to protect the cyber environment and organization and assets."

It is worth noting that definition #18 is part of an more exhaustive definition text by the International Telecommunication Union (2008), but comes out top due to its relative conciseness. On the flip side, we note that brevity is not key to a representative definition (in context of the pool of our authoritative definitions) as illustrated by the trailing definitions #16, #17 and #28. These are very concise but do not have sufficient descriptive depth to capture the meaning of cybersecurity; both objectively, as shown in the comparison, as well as subjectively (although this leaves plenty of room for argument).

Table 2.
Top 5 most representative definitions

ID	source	title	SimScore
18	Republic of South Africa	Cybersecurity Policy of South Africa	0.434
11	French Network and Information Security Agency	Information systems defence and security France's strategy	0.426
21	Spanish Cyber Security Institute	National Cyber Security, a commitment for everybody	0.409
3	International Telecommunication Union	Series X: Data Networks, Open System Communications and Security	0.407
15	New Zealand Government	New Zealand's Cyber Security Strategy	0.405

It is important to point out we didn't identify this to be the most relevant definition through expert opinion but through unbiased similarity analysis based on an authoritative set of definitions. Definition #18 best captures the essence of all authoritative definitions combined.

6. TOWARDS AN IMPROVED DEFINITION

After identifying the most representative definitions for 'cyber security' as described in the previous section, the next step was to try and construct an improved definition. The new definition would then be measured under the same conditions to compare similarity scores. Using KH Coder (Higuchi, 2015) we investigated the previously mentioned top five definitions (18, 11, 21, 3, 15) with the assumption that they contain the most relevant attributes in the overall definition pool. To establish the key underlying concepts needed to craft an improved definition, we used co-occurrence network analysis (Rice & Danowski, 1993). In textual analysis, co-occurrence networks show words with similar appearance patterns and as such with high degrees of co-occurrence. The approach is based on the idea that a word's meaning is related to the concepts to which it is connected. It also has the benefit that no coder bias is introduced other than to determine which words are examined (Ryan & Bernard, 2003). However, applying the function on our definition set, even though already limited to five paragraphs and with minimum spanning tree applied, proved to produce a very crowded output difficult to navigate. By filtering for term frequency (TF ≥ 2) when producing the co-occurrence network graph, we were able to reduce the information presented to a (human) manageable level while preserving important context.

Figure 8 shows the minimum spanning tree (MST) network graph model with 32 nodes and 25 edges extracted. The graph presents an at a glance a view of the underlying concepts inherent to the words used in the definition set. In addition to the minimum spanning tree, we have added community detection to further emphasize connected components. The node size illustrates the term frequency, and detected communities are highlighted in different colors. Based on the dataset, we found that the 'random walk' or 'walktrap' algorithm (Pons & Latapy, 2005) provided the subjectively best community detection approach. Combined with MST, it aids in understanding not only the key concepts but also how words group into communities and which communities are closer to each other (signified by dotted lines).

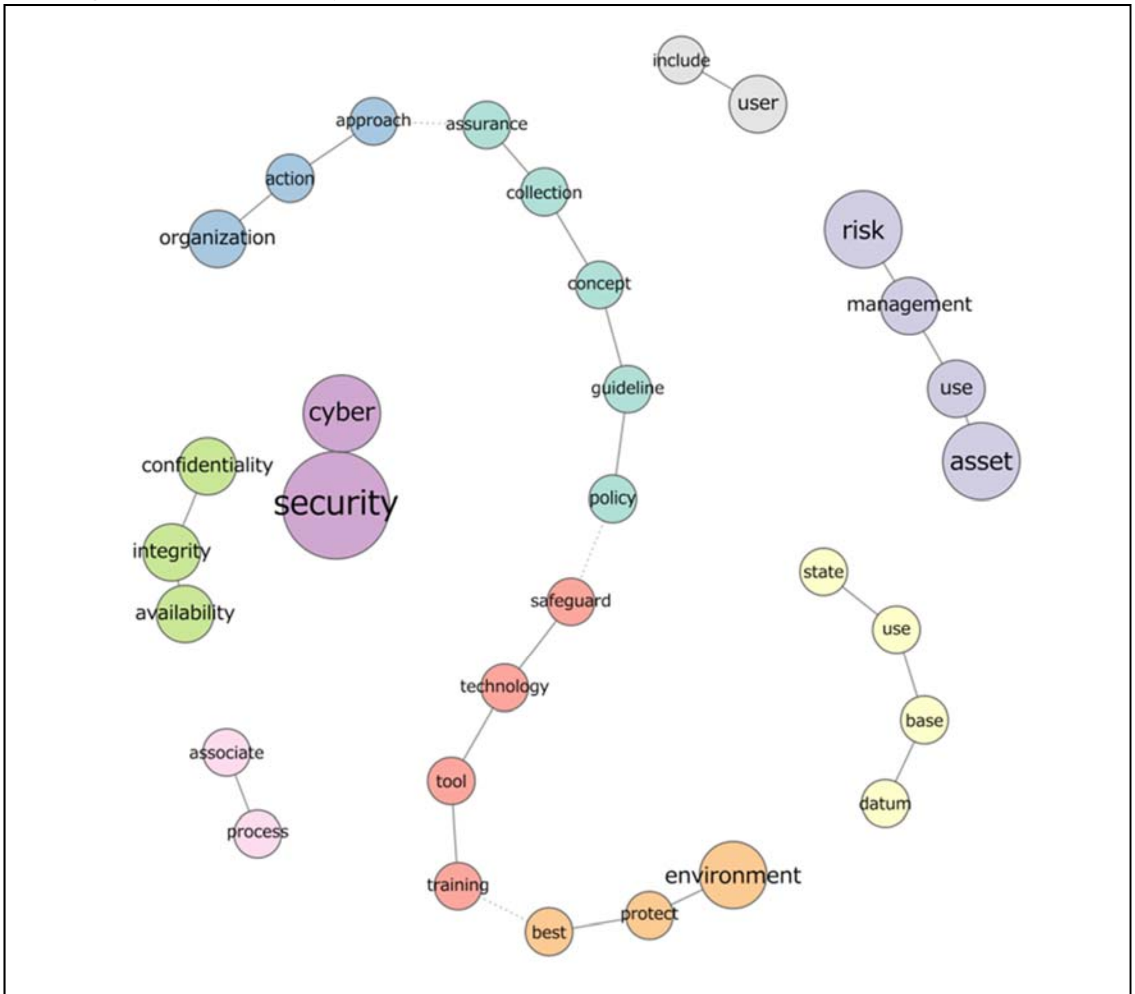


Figure 8. Top 5 correspondence analysis (TF>2)

With the key components extracted, we were in a position to create our own proposal for an improved definition. Through several iterations of manual sentence construction using words and communities, we arrived at a definition that captures key components and respects community adhesion;

> “The approach and actions associated with security risk management processes followed by organizations and states to protect confidentiality, integrity and availability of data and assets used in cyber space. The concept includes guidelines, policies and collections of safeguards, technologies, tools and training to provide the best protection for the state of the cyber environment and its users.”

To verify that this definition is not only representative from a human reader perspective but also in terms of semantic similarity, we repeated our semantic analysis benchmarking work (Section 5), this time including our new definition (#29).

As expected, the overall results are nearly the same as previously since the methodology and configuration of the benchmark has not changed. Individual scores have changed slightly due to the new addition (29) to the corpus. The overall ranking did not change except for our proposed definition being included at the top as seen in

Table **3**.

Table 3
Top results for improved definition dataset

ID	source	title	SimScore
29	New Definition	n/a	0.465
18	Republic of South Africa	Cybersecurity Policy of South Africa	0.440
11	French Network and Information Security Agency	Information systems defence and security France's strategy	0.434
21	Spanish Cyber Security Institute	National Cyber Security, a commitment for everybody	0.416
3	International Telecommunication Union	Series X: Data networks, open system communications and security	0.412
15	New Zealand Government	New Zealand's Cyber Security Strategy	0.409

7. STUDY LIMITATIONS AND CHALLENGES TO VALIDITY

In the previous section, we proposed a new definition for 'cyber security' which tops the ranking of most relevant definitions among authoritative sources. However, as with many similar research exercises, there is no claim to completeness or infallibility of our work. Our study is affected by limitations inherent to literature reviews as described by Kitchenham and Charters (2007) which includes limitations on search comprehensiveness and material selection. To mitigate this weakness, forward and backward reference checking was conducted on key publications to discover potentially relevant sources. Regardless, it is possible that our efforts missed sources which we would have otherwise considered authoritative and relevant (although the number of definitions covered in this study should ensure relevance irrespective). Another inherent limitation to literature reviews is the language barrier, as this work only covered definitions provided in English.

Although the study has achieved its objective of creating a representative definition for 'cyber security,' our approach for creating the definition is limited by manual sentence generation constrains. It is possible that an automated approach, iterating all possible combinations of our nodes and communities leveraging natural language generation (Sauper & Barzilay, 2009), would have produced another, perhaps more relevant definition. This was beyond the scope of this paper but will be considered for future work.

Lastly, considering the pace at which social communities create, adopt and modify their understanding of developing areas such as 'cyber' and 'cyberspace,' our definition is representative at the time of the research. It is expected that this definition will become less fitting or relevant as social, political, and technological developments in this space progress. Nonetheless, our proposed model for evaluating definitions will prove useful and remain relevant in the future.

8. CONCLUSION AND RESEARCH QUESTIONS

For this research, we set out to analyze the landscape of authoritative sources defining the term 'cyber security.' As part of this work, we conducted a semi-systematic literature review identifying relevant sources. Through our efforts as outlined in section 3, we found 28 authoritative sources fulfilling our inclusion criteria and were included for further analysis in context of our research questions. This not only provided the fundament to answer our research questions, but also contributed the most exhaustive set of authoritative sources for further research in this field. We found the majority of definition sources to be related to governmental institutions with several additional relevant sources from industry and the academic sector (RQ1). Our review of primary sources unveiled a clear lack of congruence across the sources as to the meaning and scope of the term. Even contradictory claims in regard to scope were identified for several primary studies (RQ2). To better understand the differences in the definition set (RQ2) and to identify the most relevant definition (RQ3), we applied basic (section 4) and advanced (section 5) semantic similarity analysis methods to the data set. To our knowledge, this is the first endeavor to make use of this novel and non-biased approach to identify the most representative definition in a set of definitions (for 'cyber security'). We were able to show that the definition contributed by the Republic of South Africa (2010) achieved the highest similarity score and as such was the most representative definition of 'cyber security' under the conditions of this work. To answer our final research question (RQ4), we conducted further analysis on the data set making use of co-occurrence, semantic networks, and community detection methods. By isolating key components and communities in the definition set, we produced an improved definition for 'cyber security' (section 6). Our new definition was shown to be the new most representative definition following the same methodology discussed in the paper. While we recognize the potential for further improvement of this approach (section 7), we believe that the methodology and the improved definition is a noteworthy contribution to the field.

REFERENCES

Androutsopoulos, I., & Malakasiotis, P. (2010). A survey of paraphrasing and textual entailment methods. *J. Artif. Int. Res., 38*(1), 135-187.

Barzilay, M. (2013, 2013-08-05). A simple definition of cybersecurity. Retrieved from http://www.isaca.org/Knowledge-Center/Blog/Lists/Posts/Post.aspx?ID=296

Baylon, C. (2014). *Challenges at the Intersection of Cyber Security and Space Security: Country and International Institution Perspectives.* Retrieved from London: http://www.chathamhouse.org/publication/challenges-intersection-cyber-security-and-space-security-country-and-international

Choi, H., & Varian, H. A. L. (2012). Predicting the Present with Google Trends. *Economic Record, 88,* 2-9. doi:10.1111/j.1475-4932.2012.00809.x

Congressional Research Service. (2014). *Cybersecurity Issues and Challenges: In Brief.* (R43831). Retrieved from https://www.fas.org/sgp/crs/misc/R43831.pdf.

Corley, C., & Mihalcea, R. (2005). *Measuring the semantic similarity of texts.* Paper presented at the Proceedings of the ACL Workshop on Empirical Modeling of Semantic Equivalence and Entailment, Ann Arbor, Michigan.

Cornell University. (2016). Critically Analyzing Information Sources: Critical Appraisal and Analysis. Retrieved from http://guides.library.cornell.edu/c.php?g=31866&p=201757

Couto, F. M., Silva, M. J., & Coutinho, P. M. (2007). Measuring semantic similarity between Gene Ontology terms. *Data & Knowledge Engineering, 61*(1), 137-152. doi:http://dx.doi.org/10.1016/j.datak.2006.05.003

Craigen, D., Diakun-Thibault, N., & Purse, R. (2014). Defining Cybersecurity. *Technology Innovation Management Review, 4*(10).

Creasey, J. (2013). Cyber Security Incident Response Guide, 56. Retrieved from http://www.crest-approved.org/guidance-and-standards/cyber-security-incident-response-guide/index.html

De Marneffe, M.-C., MacCartney, B., & Manning, C. D. (2006). *Generating typed dependency parses from phrase structure parses.* Paper presented at the Proceedings of LREC.

Eig, L. M. (2011). *Statutory Interpretation: General Principles and Recent Trends* (97-589). Retrieved from https://fas.org/sgp/crs/misc/97-589.pdf

Falessi, N., Gavrila, R., Klejnstrup Ritter, M., & Moulinos, K. (2012). *Practical Guide on Development and Execution.* Retrieved from Heraklion: http://www.enisa.europa.eu/activities/Resilience-and-CIIP/national-cyber-security-strategies-ncsss/national-cyber-security-strategies-an-implementation-guide

Giles, K., & Hagestad, W. (2013, 4-7 June 2013). *Divided by a common language: Cyber definitions in Chinese, Russian and English.* Paper presented at the Cyber Conflict (CyCon), 2013 5th International Conference on.

Godwin III, J. B., Kulpin, A., Rauscher, K. F., & Yaschenko, V. (2014). *Critical Terminology Foundations 2.* Retrieved from New York: http://www.ewi.info/idea/critical-terminology-foundations-2

Gomaa, W. H., & Fahmy, A. A. (2013). A survey of text similarity approaches. *International Journal of Computer Applications, 68*(13), 13-18.

Government of Montenegro. (2013). *National Cyber Security Strategy for Montenegro 2013-2017.* Podgorica Retrieved from http://www.mid.gov.me/ResourceManager/FileDownload.aspx?rid=165416&rType=2&file=Cyber%20Security%20Strategy%20for%20Montenegro.pdf.

Graesser, A. C., Olney, A., Haynes, B. C., & Chipman, P. (2005). AutoTutor: A Cognitive System That Simulates a Tutor Through Mixed-Initiative Dialogue. *Cognitive systems: Human cognitive models in systems design*, 177.

Hearst, M. A. (1999). *Untangling text data mining.* Paper presented at the Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics, College Park, Maryland.

Higuchi, K. (2015). KH_Coder (Version 2). Retrieved from http://khc.sourceforge.net/

International Telecommunication Union. (2008). Overview of cybersecurity *SERIES X: DATA NETWORKS, OPEN SYSTEM COMMUNICATIONS AND SECURITY* (pp. 64).

Internet Society. (2012). Some Perspectives on Cybersecurity, 22. Retrieved from Internet Society website: http://www.internetsociety.org/doc/some-perspectives-cybersecurity-2012

ISACA. (2014). *European Cybersecurity Implementation: Overview.* Retrieved from Rolling Meadows: http://www.isaca.org/Knowledge-Center/Research/ResearchDeliverables/Pages/European-Cybersecurity-Implementation-Series.aspx

Jiang, J. J., & Conrath, D. W. (1997). *Semantic similarity based on corpus statistics and lexical taxonomy.* Paper presented at the In the Proceedings of ROCLING X, Taiwan.

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering.* Retrieved from http://www.dur.ac.uk/ebse/resources/Systematic-reviews-5-8.pdf

Leacock, C., Miller, G. A., & Chodorow, M. (1998). Using corpus statistics and WordNet relations for sense identification. *Computational Linguistics, 24*(1), 147-165.

Lee, M. C. (2011). A novel sentence similarity measure for semantic-based expert systems. *Expert Systems with Applications, 38*(5), 6392-6399. doi:http://dx.doi.org/10.1016/j.eswa.2010.10.043

Lin, D. (1998). *An information-theoretic definition of similarity.* Paper presented at the 15th International Conference on Machine Learning, Madison, WI.

Lintean, M. C. (2011). *Measuring semantic similarity: representations and methods.* The University of Memphis.

Luiijf, E., Besseling, K., & de Graaf, P. (2013). Nineteen national cyber security strategies. *International Journal of Critical Infrastructures, 9*(1), 3-31. doi:10.1504/IJCIS.2013.051608

Mäntylä, M. V., Adams, B., Khomh, F., Engström, E., & Petersen, K. (2014). On rapid releases and software testing: a case study and a semi-systematic literature review. *Empirical Software Engineering, 20*(5), 1384-1425. doi:10.1007/s10664-014-9338-4

Martin, D. I., & Berry, M. W. (2007). Mathematical foundations behind latent semantic analysis. *Handbook of latent semantic analysis*, 35-56.

Meyer, D., Hornik, K., & Feinerer, I. (2008). Text mining infrastructure in R. *Journal of Statistical Software, 25*(5), 1-54.

Miller, G. A. (1995). WordNet: a lexical database for English. *Commun. ACM, 38*(11), 39-41. doi:10.1145/219717.219748

Nakov, P., Popova, A., & Mateev, P. (2001). Weight functions impact on LSA performance. *EuroConference RANLP*, 187-193.

Pons, P., & Latapy, M. (2005). Computing Communities in Large Networks Using Random Walks. In p. Yolum, T. Güngör, F. Gürgen, & C. Özturan (Eds.), *Computer and Information Sciences - ISCIS 2005* (Vol. 3733, pp. 284-293): Springer Berlin Heidelberg.

Porter, M. F. (1997). An algorithm for suffix stripping. In J. Karen Sparck & W. Peter (Eds.), *Readings in information retrieval* (pp. 313-316): Morgan Kaufmann Publishers Inc.

Republic of South Africa. (2010). *Cybersecurity Policy of South Africa.* Pretoria.

Rice, R. E., & Danowski, J. A. (1993). Is It Really Just Like a Fancy Answering Machine? Comparing Semantic Networks of Different Types of Voice Mail Users. *Journal of Business Communication, 30*(4), 369-397. doi:10.1177/002194369303000401

Rus, V. (2014). *Opportunities and Challenges in Semantic Similarity.* Paper presented at the 2014.

Rus, V., & Lintean, M. (2012). *A comparison of greedy and optimal assessment of natural language student input using word-to-word similarity metrics.* Paper presented at the Proceedings of the Seventh Workshop on Building Educational Applications Using NLP, Montreal, Canada.

Rus, V., Lintean, M., Moldovan, C., Baggett, W., Niraula, N., & Morgan, B. (2012). *The similar corpus: A resource to foster the qualitative understanding of semantic similarity of texts.* Paper presented at the Semantic Relations II: Enhancing Resources and Applications, The 8th Language Resources and Evaluation Conference (LREC 2012), May.

Rus, V., Lintean, M. C., Banjade, R., Niraula, N. B., & Stefanescu, D. (2013). *SEMILAR: The Semantic Similarity Toolkit.* Paper presented at the ACL (Conference System Demonstrations).

Ryan, G. W., & Bernard, H. R. (2003). Techniques to Identify Themes. *Field Methods, 15*(1), 85-109. doi:10.1177/1525822x02239569

Salton, G. (1963). Associative Document Retrieval Techniques Using Bibliographic Information. *J. ACM, 10*(4), 440-457. doi:10.1145/321186.321188

Sauper, C., & Barzilay, R. (2009). *Automatically generating Wikipedia articles: a structure-aware approach.* Paper presented at the Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1 - Volume 1, Suntec, Singapore.

Sowell, T. (2014). *Basic Economics* (5th ed.). New York: Basic Books.

Sparck Jones, K. (1972). A statistical interpretatoin of term specificity and its application in retrieval. *Journal of Documentation, 28*(1), 11-21. doi:doi:10.1108/eb026526

Stubley, D. (2013, 2013-06-07). What is Cyber Security? Retrieved from https://www.7elements.co.uk/resources/blog/what-is-cyber-security/

The White House. (2009). National Cybersecurity Awareness Month, 2009 [Press release]. Retrieved from https://www.whitehouse.gov/the_press_office/Presidential-Proclamation-National-Cybersecurity-Awareness-Month/

Walls, A., Perkins, E., & Weiss, J. (2013). Definition: Cybersecurity, 5. Retrieved from Gartner.com website: https://www.gartner.com/doc/2510116/definition-cybersecurity

Wamala, F. (2011). *ITU National Cybersecurity Strategy Guide*. Retrieved from Geneva: http://www.itu.int/ITU-D/cyb/cybersecurity/docs/itu-national-cybersecurity-guide.pdf

Weiss, S., Indurkhya, N., Zhang, T., & Damerau, F. (2004). *Text Mining: Predictive Methods for Analyzing Unstructured Information*: SpringerVerlag.

Wu, Z., & Palmer, M. (1994). *Verbs semantics and lexical selection*. Paper presented at the Proceedings of the 32nd annual meeting on Association for Computational Linguistics, Las Cruces, New Mexico.

Yuhua, L., Bandar, Z. A., & McLean, D. (2003). An approach for measuring semantic similarity between words using multiple information sources. *Knowledge and Data Engineering, IEEE Transactions on, 15*(4), 871-882. doi:10.1109/TKDE.2003.1209005

Yuhua, L., McLean, D., Bandar, Z. A., O'Shea, J. D., & Crockett, K. (2006). Sentence similarity based on semantic nets and corpus statistics. *Knowledge and Data Engineering, IEEE Transactions on, 18*(8), 1138-1150. doi:10.1109/TKDE.2006.130

APPENDIX A

Table 4
Complete overview of definition sources

ID	Source	Title	Year
1	Committee on National Security Systems	National Information Assurance (IA) Glossary	2009
2	National Initiative for Cybersecurity Careers and Studies,	Explore Terms: A Glossary of Common Cybersecurity Terminology	n/a
3	International Telecommunication Union	Series X: Data Networks, Open System Communications and Security	2008
4	Gartner	Definition: Cybersecurity	2013
5	The Institution of Engineering and Technology	Resilience and Cyber Security of Technology in the Built Environment	2013
6	British Standards Institute	Guidelines for cybersecurity	2012
7	Australian Government	Cyber Security Strategy	2009
8	Federal Chancellery of the Republic of Austria	Austrian Cyber Security Strategy	2013
9	Government of Belgium	Cyber Security Strategy	2012
10	Government of Finland	Finland's Cyber Security Strategy	2013
11	French Network and Information Security Agency	Information systems defence and security France’s strategy	2011
12	Federal Ministry of the Interior	Cyber Security Strategy for Germany	2011
13	Government of Hungary	National Cyber Security Strategy of Hungary	2013
14	The Netherlands, Ministry of Security and Justice	The National Cyber Security Strategy (NCSS) 2	2013
15	New Zealand Government	New Zealand’s Cyber Security Strategy	2011
16	Norwegian Ministries	Cyber Security Strategy for Norway	2012
17	Kingdom of Saudi Arabia	Developing National Information Security Strategy for the Kingdom of Saudi Arabia	2011
18	Republic of South Africa	Cybersecurity Policy of South Africa	2010
19	Republic of Turkey	National Cyber Security Strategy and 2013-2014 Action Plan	n/a
20	National Institute of Standards and Technology	Framework for Improving Critical Infrastructure Cybersecurity	2014
21	Spanish Cyber Security Institute	National Cyber Security, a commitment for everybody	2012

ID	Source	Title	Year
22	Republic of Poland	Cyberspace Protection Policy of The Republic of Poland	2013
23	Government of Jamaica	National Cyber Security Strategy	2015
24	Craigen, Dan Diakun-Thibault, Nadia Purse, Randy	Defining Cybersecurity	2014
25	Merriam-Webster	Definition of Cybersecurity	2015
26	Oxford Dictionary	Definition of Cybersecurity	2015
27	Amoroso, Edward	Cyber Security	2007
28	EastWest Institute	Critical Terminology Foundations 2	2014
29	New Definition		2016

[i] https://www.enisa.europa.eu/activities/Resilience-and-CIIP/national-cyber-security-strategies-ncsss

[ii] https://ccdcoe.org/strategies-policies.html

[iii] Attributed to Edward Deming