

# Computational Text Analysis of A Scientific Resilience Management Corpus: Environmental Insights and Implications

J. Nassour<sup>1</sup>, D. Leykin<sup>2,3</sup>, M. Elhadad<sup>1</sup>, and O. Cohen<sup>2,4,5\*</sup>

<sup>1</sup> Computer Science Department, Faculty of Natural Sciences, Ben-Gurion University of the Negev, POB 635, Israel

<sup>2</sup> PREPARED Center for Emergency Response Research, Ben-Gurion University of the Negev, POB 635, Israel

<sup>3</sup> The Community Stress Prevention Centre (CSPC), P.O. Box 797, Kiryat-Shmona 11016, Israel

<sup>4</sup> Nursing Department, Recanati School for Community Health Professions, Faculty of Health Sciences, Ben-Gurion University of the Negev, P.O. Box 653, Beer-Sheva 84105, Israel

<sup>5</sup> Masters' program in Emergency Medicine, School of Public Health, Faculty of Health Sciences, Ben-Gurion University of the Negev, P.O. Box 653, Beer-Sheva 84105, Israel

Received 04 June 2018; revised 26 November 2018; accepted 12 July 2019; published online 18 November 2019

**ABSTRACT.** Resilience is a multifaceted concept describing the ability to cope with change or disruption. Its importance in the era of emergency preparedness and response, combined with its multidisciplinary attributes, have led researchers to study similarities and differences in the meaning of resilience across various fields. A systematic literature review, conducted in the field of resilience management by the DARWIN project, yielded a scientific corpus of 419 articles. In the present study, automated text-analysis approaches were used to investigate this corpus and generate insights, aiming at understanding resilience management. Three complementary computational analyses were employed: (a) topic modeling to understand the different topics or fields discussed in the articles; (b) concept maps to provide a synthetic view of key concepts in the domain and their relations; (c) psycho-linguistic analysis to identify significant psychological categories addressed in the corpus. The topic model identified four key topics: Environmental/Socioecological aspects, Organizational/Operational aspects, Health, and Infrastructure/Resource Management. The concept map recognized concepts at a finer granularity level and depicted them into five main clusters with relations between them, reflecting key dimensions leading to resilience management. The psycho-linguistic analysis highlighted the importance of psychological processes within resilience management. This study identified important aspects that need to be addressed when designing resilience management frameworks, such as rehabilitation period and the role of public.

**Keywords:** concept maps, LIWC, NLP tools, resilience, resilience management, topic modeling

## 1. Introduction

Resilience is a concept borrowed from the exact sciences, and has been used in ecology since Holling introduced it in his thesis in 1973. Thereafter, the notion of resilience has entered various worlds of content, describing the capacity of systems, individuals, infrastructures, organizations, or communities to cope with situations of change and crisis (Norris et al., 2008; Magis, 2010; Steiner and Atterton, 2014; Bonanno et al., 2015; Wilson, 2017). Furthermore, it was implemented in national and international frameworks for emergency coping plans (e.g., Cabinet Office, 2011; UNISDR, 2015). Numerous publications are devoted to resilience, and a Google Scholar search of the word yields approximately 1,850,000 results, with the number of publications increasing annually.

As a concept, resilience is structurally complex, with both

physical and perceptual components (Cutter, 2016). Its importance and complexity, combined with its multidisciplinary attributes, have led researchers to study similarities and differences in the meaning of resilience across various fields (Southwick et al., 2014; Hosseini et al., 2016). While experts from each field have a different perception of the meaning of resilience, they all see it as a factor that mitigates the impact of change or an emergency situation, and shortens the duration of that impact (Castleden et al., 2011).

One of the challenges of coping with emergency situations is the management of resilience in a time of change and unexpected situations so as to promote continuity of different aspects (Caralli et al., 2010; Linkov, et al., 2014). Walker et al. (2002) claim that resilience management aims at preventing a system from moving out of configuration. Resilience management affects all phases of the emergency cycle from pre-event period, to during the crisis, and including the rehabilitation phase (McManus et al., 2007). Hence, resilience management must consider a wide range of activities, including situation awareness, response and preparedness planning, assessment of resources beyond needs, implementation activities, and adaptation

\* Corresponding author. Tel.: +972 86478430; fax: +972 86472136.  
E-mail address: odayac@bgu.ac.il (O. Cohen).

of actions in the face of the current situations (Sengupta, 2006).

The importance of resilience as a component of emergency preparedness and response, and the growing number of emergency situations worldwide, were the impetus for the DARWIN project (Branlat et al., 2017; Herrera et al., 2017). DARWIN (<https://www.h2020darwin.eu/>), which originated within the Horizon 2020 program, was conceived with the intention of improving response to expected and unexpected crises affecting critical infrastructures and social structures. The aim of this project was to develop state-of-the-art resilience guidelines and innovative training modules for crisis management, both natural and man-made (DARWIN, 2015; Adini et al., 2017). The project included nine participating organizations from different countries, with each organization contributing its experience and knowledge to maximize the impact of the project outcomes. The DARWIN project focused on two fields: healthcare and air-traffic management. However, its basic approach was to develop and provide general guidelines for a variety of emergency situations, assisting different organizations develop and enhance their resilience in the context of crisis management (DARWIN, 2017).

The first stage of the DARWIN project was to understand different perspectives of resilience management, and to identify the concepts, theories and practices of resilience management as reflected in the professional literature. To achieve this goal, a four-step vast systematic literature review was conducted. The review resulted in a corpus of 419 articles, which were found relevant for further synthesis and analysis through full paper review. This corpus constitutes the foundation of the data used for the present analysis.

The multi-disciplinary nature of the “resilience” concept leads to two difficulties in making professional literature accessible for researchers: (a) the sheer volume; (b) the variability in terminology, methodology, and rhetorical structure of the articles issued from different disciplines makes it challenging to create a synthesis that can be re-used by researchers in different fields.

According to Blei (2012), computational text analysis enables gaining insights, and discovering main themes that exist in a corpus of unstructured references. Text analysis is a task that belongs to the field of Natural language processing (NLP). NLP addresses the problems of automatically extracting important information from text, in addition to understanding and generating text. It started in the 1950s as the intersection of linguistics and artificial intelligence. In the 1990s, NLP started utilizing statistics-based machine learning algorithms for these tasks (Nadkarni et al., 2011).

Recent progress in Natural Language Processing (NLP) and Text Mining enables the practice of “distant reading” – i.e., extracting key concepts and their relations in a corpus, analyzing trends over time, identifying terminological variants and producing usable synthetic visualizations of the contents (Moretti, 2013). Different research fields (e.g., Digital Humanities, Marketing, Transportation Research, Bioinformatics) have already leveraged this methodology in their research process (Amado et al., 2017; Hahn et al., 2017; Sun and Yin, 2017).

Due to their non-manual and statistical nature, NLP tools can assist researchers cover a large spectrum of the literature and explore trends in various fields based on published literature.

The relationship between environment and resilience is well established in the literature (e.g., Magis et al., 2010; Lizarralde et al., 2015; Cerè et al., 2017). According to Marchese et al. (2018), implementation of resilience frameworks provides benefits to both humans and the environment. This study introduces three specific NLP tools – topic modeling, concept maps, and psycho-linguistic analysis – for the analysis of a multidisciplinary corpus on resilience management with the objective of obtaining both general and environmental insights towards the design of a comprehensive framework for resilience management.

## 2. Materials and Methods

### 2.1. Data

The DARWIN corpus (DARWIN, 2015) is utilized for the analysis performed in this paper. This corpus was the result of a four-step systematic literature review conducted using Kitchenham’s methods (2004): planning, conducting, extracting/synthesizing data, and reporting. Based on inclusion criteria of relevance and quality, the systematic review yielded 419 articles that were found to use scientific methodology in their investigation of resilience management. The selected papers were published between 1998 and 2015, representing a wide range of domains, including community/social, ecology, infrastructure, organizational, and economics. To get an impression of the articles, as well as a full description of the process and the results, refer to the first work package of the DARWIN project (DARWIN, 2015).

The dataset used for this study contains the full text of 392 articles, out of the original 419. The rest of the articles were inaccessible in electronic form. The corpus is mostly in PDF format but some of the articles are in HTML. Text was extracted from PDFs using Dr. Inventor (Ronzano and Saggion, 2015), a framework designed to facilitate analysis of scientific articles. In the case of HTMLs, manually crafted patterns over HTML tags and their attributes were employed. PDF extractions were manually fixed in cases where only a small part of the text was extracted.

To improve the analysis of the text, pre-processing steps were performed to clean the data. The steps were: tokenization using Stanford PTBTokenizer (CoreNLP, 2016), conversion into lower case, removal of stop words and special characters, and stemming using NLTK’s Porter stemmer (Bird, 2006). Collectively, these steps have the benefit of reducing the number of distinct words observed in the dataset, by normalizing variants into a single canonical form.

### 2.2. Topic Modeling

Topic Modeling is an unsupervised statistical method for text analysis (Blei, 2012). It identifies the topics that a document discusses from a shared set of topics among a collection of documents. The topics are learnt automatically from the pro-

vided data with no a priori assumptions regarding the disciplines associated with the documents. Topics, in this context, are defined by lists of words that are extracted automatically from a collection of documents. The lists of words identified by the algorithm are affected by the distribution of words in the document collection. The topics are not predefined. They are discovered by the statistical analysis. Topic modeling algorithms only generate lists of words, not necessarily disjoint, without providing a name for each list. The labeling of each word list into a topic is done manually in a post-processing stage. A survey of topic modeling is provided in (Alghamdi and Alfalqi, 2015).

Non-negative Matrix Factorization (NMF) is one of the algorithms that perform topic modeling. It is an unsupervised statistical method – i.e., it does not require a priori annotations of the text corpus (Lin, 2007; Cichocki and Phan, 2009). NMF takes as input the number of desired topics and relies on linear algebra to factorize a document-term probability matrix representation of the dataset into two smaller matrices. The text corpus is encoded as a document-term matrix in the following way: each row is a document represented by the frequencies of the corpus words in the document, while each column corresponds to a distinct word type in the vocabulary. After the algorithm completes, this matrix is transformed into two smaller, informative matrices: one has a topic for each row, and a term for each column. It describes the composition of the topics identified in the corpus. The other matrix has one document for each row, and a topic for each column. It describes which topics are expressed in a specific document.

### 2.3. Concept Maps

A concept map is a graph connecting concepts (nodes) and relations (edges) between them. It provides a summary of a domain. TechKnAcq (Gordon et al., 2017) is a system for automatically generating reading lists in a domain. It selects documents based on their pedagogical value to the learner, and orders them using the structure of concepts in the domain. As a subtask, TechKnAcq builds a conceptual map of the given domain. Concepts are modeled as topics, extracted using a topic model algorithm and referred to using the top 3 ranking words that appear in the topic. Relations are predicted using information-theoretic measures (e.g., co-occurrence, cross-entropy, Jaccard similarity coefficient between the top-20 words or phrases associated with each concept, Jaccard similarity coefficient between the sets of the top-k documents associated with each concept). TechKnAcq was applied to the corpus and incorporated into this analysis.

The concept map extraction tool identifies “concepts” at a much lower granularity than the topic model analysis. For comparison, concept map extraction was applied with a parameter of fifty concepts, while for identifying the key topics, NMF was executed with a parameter of four topics.

### 2.4. Psycho-Linguistic Analysis Using LIWC

Linguistic Inquiry and Word Count (LIWC; version 2015)

is a text-analysis software that counts words in psychologically meaningful categories. LIWC can shed light on the different psychological categories addressed in a corpus (Tausczik and Pennebaker, 2010; Pennebaker et al, 2015). The LIWC dictionary is composed of about 4,500 predefined words and word stems. The dictionary was created by experts to cover psychological processes as described by the different disciplines. Each entry defines one or more word categories or sub-dictionaries. Currently, there are 91 categories of words, and four prime categories: Summary Language Variables, Linguistic Dimensions, Other Grammar, Psychological Processes, and Punctuation. Psychological Processes is further divided into ten subcategories such as social, cognitive, perceptual, and biological processes (Pennebaker et al., 2015). LIWC provides the means of the frequencies of words in each category ranging from 0 to 100.

In this study LIWC was conducted on the four clusters of articles corresponding to the four topics identified by the topic model analysis, followed by One-way analysis of variance (one-way ANOVA) on the results, to find significant differences between the distributions of word categories across topics. Means (M), Standard Deviations (SD), One-Way ANOVA *F*-statistic (*F*) and significance (*p*-values) are reported for each of the categories.

## 3. Results

The study applies three NLP tools for analyzing a corpus consisting of 392 systematically collected professional publications dealing with resilience management. The following is a report of the results presented by the tools.

### 3.1. Topic Modeling with NMF

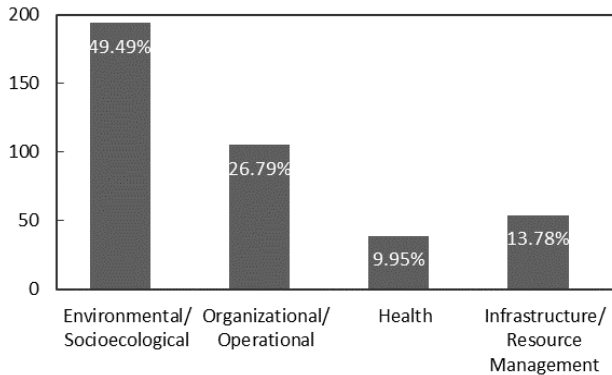
A topic model with four topics was created using NMF. The number of topics was decided empirically after several trials and consultations with experts in the field of resilience, where different numbers from three to seven were tested. The number four gave the most coherent results. The algorithm was applied several times to ensure its stability. The names of the topics were manually assigned according to the generated word-lists. The results show that the articles discuss the following topics: Environmental/Socioecological aspects of resilience management (e.g., hazard, climate, social, vulnerability); Organizational/Operational aspects (e.g., organization, operation, accident); Health (e.g., health, hospital, medical); and Infrastructure/Resource Management (e.g., Critical Infrastructure, failure, water).

Table 1 lists the top-20 words associated with each topic, and Figure 1 shows the distribution of each topic as a key topic in the dataset. Each article can mention multiple topics, and the NMF analysis indicates the strength of association between an article and a topic. When considering only the top-topic associated with each article, almost half of the articles, 49% focus on Environmental/Socioecological issues, 27% focus on Organizational/Operational issues, 14% focus on Infrastructure/Resource Management, and 10% of the articles discuss Health issues. In general, 8% of the articles discuss a single topic. 33%

discuss exactly 2 topics, 7% of them discuss the topics equally, 46% discuss exactly 3 topics, and 0.55% of them discuss the topics equally. Only 13% of the articles discuss four topics, while only 1% discuss a fourth topic with a probability of at least 10%. This indicates that only few publications have taken into account all these topics. Figure 2 displays a more detailed distribution of the topics in the dataset reflecting the number of times each topic appeared as a main, secondary, third, or fourth topic. All of the topics with the exception of Environmental/Socioecological are addressed more as a sec-ondary topic than a key topic.

**Table 1.** The 4 Topics Received by the Topic model, Together with Their Top-20 Words

Topic	Top-20 Words
Environmental/Socioecological	disaster, social, local, vulnerability, risk, hazard, govern, recovery, adapt, capacity, plan, city, change, flood, people, climate, nature, build, economy, resource
Organizational/Operational	organ, crisis, operation, information, safety, plan, risk, process, organization, situation, event, control, team, accident, learn, perform, incident, engineering, model, difference
Health	health, prepared, public, hospital, medical, patient, disaster, care, healthcare, coalition, depart, children, train, plan, engagement, nurse, staff, family, disease, county
Infrastructure/Resource Management	infrastructure, network, model, interdependency power, failure, CI (Critical Infrastructure), service, critic, simulation, damage, figure, disrupt, analysis, scenario, earthquake, water, transport, depend



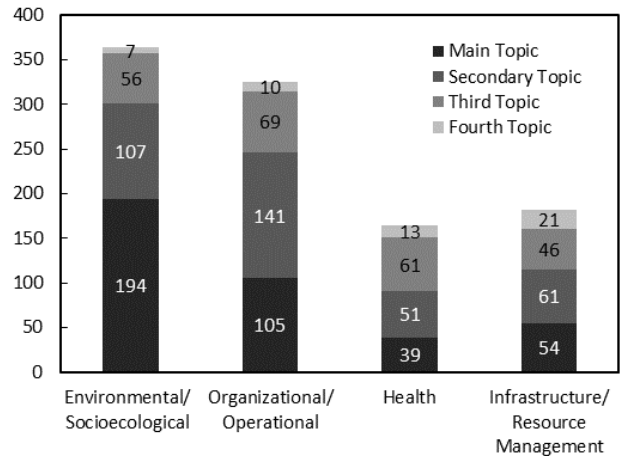
**Figure 1.** Histogram of number of articles per topic. The values inside the bins are the percentage of articles associated with each topic as its key topic.

### 3.2. Concept Maps with TechKnAcq

TechKnAcq was applied to the corpus to extract concepts and relations between them. Several trials were made, with different number of concepts, until a satisfactory map was received. The resulting concept map is shown in Figure 3, which was generated with a concept number of 50. The map displays concepts which are connected to other concepts.

Figure 3 depicts five main clusters relating to resilience management: 1) *systems preparedness* (*plan\_simul\_trains*, *system\_safeti\_oper*), 2) *emergency response plan* (*emerg\_repons\_plan*), 3) *disaster risk vulnerability* (*disas\_risk\_vulner*, *flood\_climat\_risk*), 4) *community resilience* (*commun\_resili\_disast*, *resili\_commun\_climat*), 5) *hazard risk mitigation* together with *disaster recovery plan* (*hazard\_mitig\_risk*, *recov\_eri\_paln\_disast*, *hous\_reconstruct\_recoveri*). These clusters reflect the various dimensions of the process of resilience management. Two of the clusters are related to each other through resilience indicators of the community, *resili\_indic\_commun*.

Mapping each concept to a topic from the identified topics, where possible, some relations between the topics are revealed. Organizational/Operational is related to Infrastructure/Resource Management (i.e., *system\_resili\_engin* is connected to *model\_analysi\_simul* and *system\_safeti\_oper* is connected to *service\_model\_oper*). Infrastructure/Resource Management is linked to Health (i.e., *service\_model\_oper* is connected to *hospi\_t\_care\_patient*). Organizational/Operational is related to Health through Infrastructure/Resource Management (i.e., *system\_safeti\_oper* is connected to *service\_model\_oper* which is connected to *hospital\_care\_patient*).



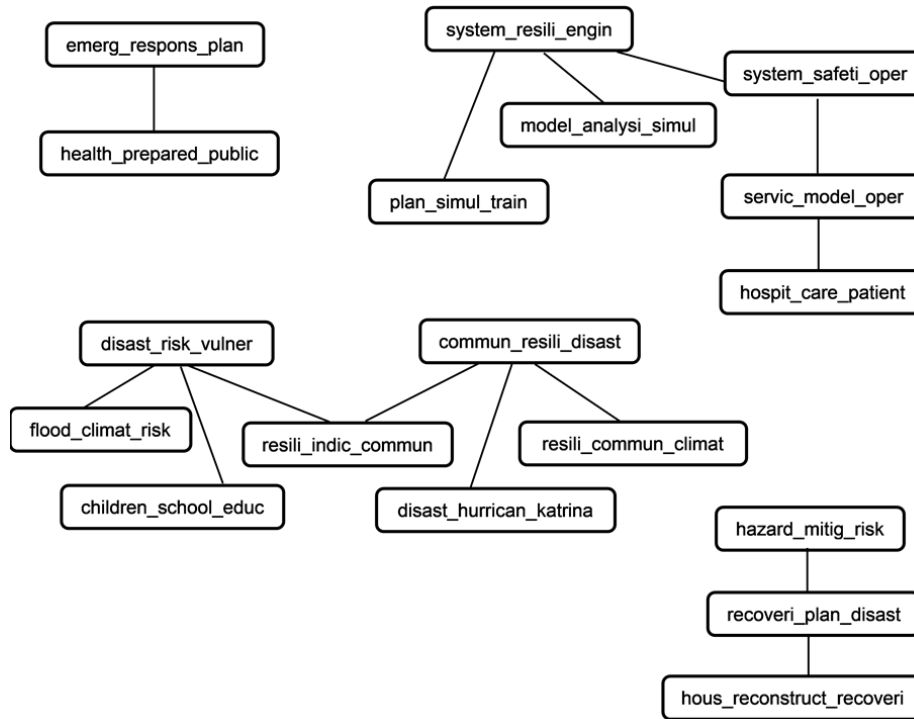
**Figure 2.** Histogram of the number of articles that address each of the extracted topics as the main topic, as the secondary topic, as the third topic, and as the fourth topic.

### 3.3. Psycho-Linguistic Analysis with LIWC

LIWC was applied on the scientific literature of each of the four topics as identified by the NMF topic modeling (see 3.1). A word cloud is employed to highlight the categories that are significant for each of the topics.

Figure 4 shows the word cloud of 47 LIWC categories, size-sorted by F-statistics derived from the one-way ANOVA analysis presented in Table A.1. Categories are associated with the topic in which they appear most frequently. The word cloud shows that each topic has several word categories that differentiate it from the other topics.

Articles belonging to Health have (as expected) a significantly higher proportion of words belonging to Health F



**Figure 3.** A concept map of the resilience domain generated by running TechKnAcq on the corpus (each concept is represented by concatenating the top-3 words that represent it).



**Figure 4.** Word Cloud of 47 LIWC categories, size-sorted by F-statistics derived from One-Way ANOVA analysis and colored by resilience topic, indicating the topic with the highest score for the specific word-category.

(3, 388) = 153.29,  $p < 0.001$  (i.e., words related to medical treatment such as clinic and pill), Core Drives and Needs (i.e., words related to affiliation, reward, risk/prevention focus), social words related to medical treatment such as clinic and pill), *Core Drives and Needs* (e.g., Friends and Family), and Positive Emotions. Infrastructure/Resource Management related articles have a significantly higher proportion of Negative Emotions, specifically those related to Sadness  $F(3, 388) = 19.23$ ,  $p < 0.001$  and Anger. Organizational/Operational articles have a significantly higher prevalence of Cognitive Processes  $F(3, 388) = 27.73$ ,  $p < 0.001$  (e.g., Insight and Causation), and content related categories, such as Work and Achievement. Environmental/Socioecological related articles had a significantly higher prevalence of categories related to Relativity (specifically Space  $F(3, 388) = 17.72$ ,  $p < 0.001$ ), in addition to categories related to personal concerns (i.e., Religion and Home), and a higher proportion of Negative Emotions.

Beyond these psychological elements, the LIWC provides information regarding language usage in the different topics. Health has significantly higher values for Clout  $F(3, 388) = 34.90$ ,  $p < 0.001$ , indicating a confident style, speaking from the perspective of high expertise, while Infrastructure/Resource Management has significantly higher values for Analytical Thinking,  $F(3, 388) = 9.20$ ,  $p < 0.001$ . Organizational/Operational is characterized by interrogative language, Interrog  $F(3, 388) = 13.27$ ,  $p < 0.001$ , directed towards understanding and describing the specifics of how and when processes are conducted.

#### 4. Discussion

The current study uses computational methods for analyzing a corpus of multidisciplinary professional publications on resilience management. The tools complement each other, with each tool giving a different perspective. To our knowledge, these methods have not been applied yet in the area of resilience management. The following is an analysis of the results received by each computational tool, followed by a summary of the main findings.

##### 4.1. Topic Modeling Analysis with NMF

Four main topics were generated using the NMF topic model: Environmental/Socioecological aspects, Organizational/Operational aspects, Health, and Infrastructure/Resource Management. The variety in the topics reflects the different content worlds involved in resilience management (see for example Cutter et al.'s dimensions of resilience (2006)). The distribution of each topic, presented in Figure 1, indicates that Environmental/Socioecological is the main topic in the corpus. This can be explained by the history of the resilience approach, and its spread from the exact sciences into the Disaster Risk Reduction (DRR) perception, starting in the ecological field (Holling, 1973; Norris et al., 2008). Furthermore, the number of environmental emergencies that require resilience management has been increasing in the last decades (Birkmann and Teichman, 2010).

The Health topic received the lowest frequency in this cor-

pus, even though expected outcomes of successful resilience management involve health factors, including reduction of losses in lives, and livelihood promotion (United Nations, 2015). This can be attributed to the appearance of Health in the corpus less as a main topic compared to other topics, indicating the importance of health aspects to resilience management (Figure 2). Figure A.2 shows that Health emerged as a major core component starting from 2008.

##### 4.2. Concept Map Analysis with TechKnAcq

According to the concept map analysis, five diverse dimensions were revealed among the resilience management process (Figure 3). The health aspect appears in two different dimensions: emergency response plans, and system engineering. Thus, it is addressed in different studies and frameworks that deal with the building of resilience management (e.g., Sendai framework for disaster risk reduction 2015 ~ 2030, 2015; Homeland Security Presidential Directive 21, 2007), in agreement with the findings of the topic modeling.

The concept map exposes the phases of the emergency cycle emphasizing their essentiality to resilience management, from the pre-emergency phase to rehabilitation. It is interesting that hazard risk mitigation was identified as a separate dimension linked to disaster recovery plan. This finding expresses the importance that the literature attributes to the rehabilitation phases. It shows the progress that has been made in recent years, from being hardly addressed to becoming an integral part in the resilience management process (Smith and Wenger, 2007). The fact that recovery plans were found to be separated from re-sponse plans reflects the attention that should be given to the rehabilitation period in the context of disaster recovery both in research and practice (Smith et al., 2018).

With respect to the association between vulnerability and community resilience dimensions, the perspectives in literature vary. Some researchers perceive these terms as antonyms (Norris et al., 2008), while others see these terms as complementary expressions (Miller, et al., 2010). According to (Cutter, et al., 2006), vulnerability is an inherent characteristic or a quality of social systems, which creates the potential for harm; while resilience demonstrates the ability to respond to and recover from crises. Both of the terms include inherent components, reflecting the characteristics of the social system, as well as other components related to an external disturbance. In accordance with Cutter et al.'s (2006) perspective, vulnerability was found to be separated from the resilience dimension in the collected corpus, indicating that they are not placed on the same continuum. The association with educational aspects (*children\_school\_educ*) in the proposed model further supports this approach.

Hurricane Katrina (*disast\_hurrican\_katrina*) was a significant emergency situation that had fundamental effects on research and practical implications on disaster preparedness and response (Norris et al., 2008). Although the exposed inhabitants at Hurricane Katrina were mostly at risk population (Myers et al., 2008; Zoraster, 2010), this event is linked in the concept map analysis to the community resilience dimension rather than to the vulnerability dimension. This can be attributed to the emphasis

that was given to the importance of community resilience during a long term event, which was conceived due to the lack of governance in Hurricane Katrina (Baker and Refsgaard, 2007); particularly since leadership and governance play a significant role in the community resilience paradigm (Cox and Perry, 2011; Castleden et al., 2011; Wilson, 2012, Berkes and Ross, 2016).

### **4.3. Psycho-Linguistic Analysis with LIWC**

The importance of the LIWC lies in that its categories were created by experts in multidisciplinary domains (Tausczik and Pennebaker, 2010; Pennebaker et al., 2015), as opposed to more automatic approaches (e.g., topic modeling). Based on the LIWC analysis (see Table A.1), death and sexuality (sexual) were seldom mentioned by the different topics. Although death is an outcome of emergency situations, resilience management literature seems to focus on mitigation and recovery rather than reporting outcomes. Moreover, the categories of preparedness (prep), time, and motion are thoroughly mentioned in the corpus, with each category having an equal frequency in the four topics. The results reveal the importance of these aspects when designing emergency response plans.

Environmental/Socioecological aspects deal with physical and sociocultural elements (Cerè et al., 2017). The literature describes natural disasters that causes damage to the physical environment of the individual (Home), and relates to cultural aspects (Religion) as emergency coping factors (Alawiyah et al., 2011).

Interestingly, the use of psycholinguistic analysis reveals that even though resilience is a term that belongs to a positive psychology stream (Seligman and Csikszentmihalyi, 2014), the corpus under study has an overall negative tone (less than 50, see Table A.1). Not surprisingly, this tone is due to the negative context surrounding resilience (Bonanno, 2004), like emergency incidents, trauma and catastrophes. Considering the emotion reported in the literature of this corpus, Environmental/Socioecological articles generally describe damage and its influences (negative emotion), while Health articles focus on positive thinking (positive emotions). A more balanced narrative in terms of sentiment, can help the manuscript sound more objective. Furthermore, enhancing the positivity of the manuscript can empower the reader (Bradley and Lang, 1999).

### **4.4. Summary**

The presented methods analyze the corpus from different complementary aspects providing a comprehensive analysis of the knowledge described in the corpus. Based on this study, the resilience management process is anchored on content from Environmental/Socioecological, Health, Infrastructure/Resource Management, as well as Organizational/Operational arenas. However, according to the results of the topic modeling only around one percent of the publications in the corpus addressed all of these elements in a substantial manner. Combining psycholinguistic analysis together with topic modeling reinforces these results since the topics have significantly different distributions in the majority of the psycholinguistic categories (see Table A.1

and Figure 4).

Community and psychological processes play a significant role in resilience (e.g., Norris et al., 2008), although they were not identified as the main topics. Words related to community appeared in the topic of Environmental/Socioecological and in the topic of Health, both as words that were extracted automatically (Table 1) and as words manually chosen by experts (Table A.1). The importance of community is further illustrated in the concept map analysis where community and social aspects appear in three of the five clusters.

Psychological aspects are highlighted mainly by the LIWC, with some psychological connotations indicated in the top 20 words of the topic of Health as well as the topic of Environmental/Socioecological aspects, such as adapt and engagement. This can be attributed to the minor role that individual resilience and psychological aspects have when considering management processes. The LIWC reveals that each topic has different psychological characteristics.

Health is mentioned both independently and in relation to other concepts, since in the past it was mainly addressed as a subtopic. Communication in resilience management is a core issue (Shittu et al., 2018). Based on the LIWC analysis, language usage may vary between the academic literature of each topic. This might be also evident in the field while managing emergency situations.

## **5. Conclusion**

This study presents natural language processing techniques, demonstrating how they can be utilized to analyze the multidisciplinary field of resilience management. Applying NLP tools on a large diverse corpus of scientific articles yielded a systematic perspective, and helped gain environmental insights and implications. Specifically, this study identified the various aspects that need to be considered when designing a comprehensive resilience management framework. The study also highlighted the importance of community and psychological processes in advancing resilience management. Throughout these aspects there is a need to relate to preparedness and time when developing response plans. Special focus and attention should be directed at themes of rehabilitation and to the recovery period. These insights may assist experts, both theoretically and practically, in better understanding, and hence better addressing environmental changes and emergencies.

### **5.1. Study Limitation and Future Studies**

This study analyzed a corpus of 392 scientific articles. The full text of the remaining 27 articles (6.4%) was inaccessible. These articles were published by 21 different journals, indicating that no specific field was omitted from the analysis (see Table A.2). The creation of the concept map is not deterministic, in the sense that re-running the creation algorithm may produce a different map. Nonetheless, the chosen representative map was reviewed by researchers in the fields of resilience and emergency preparedness and response. This study was directed to a specific aspect of resilience (resilience management) with



a limited number of publications. No attempt was made to choose the same number of articles per year (see Figure A.1). Further studies should be conducted on larger corpora to get a broader perspective of resilience, and to identify trends throughout time.

**Acknowledgments.** The authors would like to thank the group of professionals – both researchers and practitioners from the DARWIN project that conducted the systematic literature review and interview of diverse experts. In particular, we would like to note the D1.1 group members which include Woltjer R., Nevhage, B., Nilsson, S., Adini, B., Cohen, O., Aharonson-Daniel L., Goldberg A., Grøtan, T.O., Branlat, M., Moe, M Frøystad, C., and Herrera, I.; Special thanks are sent to Dr. Herrera for coordinating the DARWIN project and to Dr. Woltjer for overseeing the literature review. The research leading to the results received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 653289. Opinions expressed in this publication reflect only the authors’ view. The Agency is not responsible for any use that may be made of the information it contains.

## References

- Adini, B., Cohen, O., Eide, A.W., Nilsson, S., Aharonson-Daniel, L., and Herrera, I.A. (2017). Striving to be resilient: What concepts, approaches and practices should be incorporated in resilience management guidelines? *Technol. Forecast. Soc. Chang.*, 121, 39-49. <https://doi.org/10.1016/j.techfore.2017.01.020>
- Alawayah, T., Bell, H., Pyles, L., and Rannels, R.C. (2011). Spirituality and faith-based interventions: Pathways to disaster resilience for African American Hurricane Katrina survivors. *J. Relig. Spiri. Soc. Work: Soc. Thought*, 30(3), 294-319. <https://doi.org/10.1080/15426432.2011.587388>
- Alghamdi, R. and Alfalqi, K. (2015). A survey of topic modeling in text mining. *I. J. ACSA*, 6, 147-153. <https://doi.org/10.14569/IJACSA.2015.060121>
- Amado, A., Cortez, P., Rita, P., and Moro, S. (2017). Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis. *Eur. Res. Manage. Bus. Econ.* 24(1), 1-7. <https://doi.org/10.1016/j.iedeen.2017.06.002>
- Baker, D. and Refsgaard, K. (2007). Institutional development and scale matching in disaster response management. *Ecol. Econ.*, 63(2-3), 331-343. <https://doi.org/10.1016/j.ecolecon.2007.01.007>
- Berkes, F. and Ross, H. (2016). Panarchy and community resilience: Sustainability science and policy implications. *Environ. Sci. & Policy*, 61, 185-193. <https://doi.org/10.1016/j.envsci.2016.04.004>
- Bird, S. (2006). NLTK: the natural language toolkit. *Proc. of the COLING/ACL on Interactive presentation sessions*, Sydney, Australia, 69-72.
- Birkmann, J. and Teichman, K. (2010). Integrating disaster risk reduction and climate change adaptation: key challenges-scales, knowledge, and norms. *Sustain. Sci.*, 5(2), 171-184. <https://doi.org/10.1007/s11625-010-0108-y>
- Blei, D.M. (2012). Probabilistic topic models. *Commun. ACM.*, 55(4), 77-84. <https://doi.org/10.1145/2133806.2133826>
- Bonanno, G.A. (2004). Loss, trauma, and human resilience: have we underestimated the human capacity to thrive after extremely aversive events? *Am. Psychol.*, 59(1), 20-28. <https://doi.org/10.1037/0003-066X.59.1.20>
- Bonanno, G.A., Romero, S.A., and Klein, S.I. (2015). The temporal elements of psychological resilience: An integrative framework for the study of individuals, families, and communities. *Psychol. Inq.*, 26(2), 139-169. <https://doi.org/10.1080/1047840X.2015.992677>
- Bradley, M.M. and Lang, P.J. (1999). Affective norms for English words (ANEW): Instruction manual and affective ratings, FL, Gainesville, University of Florida, Center for Research in Psychophysiology. <https://www.uvm.edu/pdodds/teaching/courses/2009-08UVM-300/docs/others/everything/bradley1999a.pdf>
- Branlat, M., Woltjer, R., Save, L., Cohen, O., and Herrera, I. (2017). Supporting resilience management through useful guidelines. *In 7th REA Symposium* pp. 32-37.
- Cabinet Office (2011). Strategic national framework on community resilience. (United Kingdom). <https://www2.oxfordshire.gov.uk/cms/sites/default/files/folders/documents/fireandpublicsafety/emergency/StrategicNationalFramework.pdf>
- Caralli, R.A., Allen, J.H., and White, D.W. (2010). *CERT Resilience Management Model: A Maturity Model for Managing Operational Resilience*. Addison-Wesley Professional.
- Castleden, M., McKee, M., Murray, V., and Leonardi, G. (2011). Resilience thinking in health protection. *J. Public Health.*, 33(3), 369-377. <https://doi.org/10.1093/pubmed/fdr027>
- Cerè, G., Rezzgui, Y., and Zhao, W. (2017). Critical review of existing built environment resilience frameworks: directions for future research. *Int. J. Disaster Risk Reduct.*, 25, 173-189. <https://doi.org/10.1016/j.ijdrr.2017.09.018>
- Cichocki, A. and Phan, A.H. (2009). Fast local algorithms for large scale nonnegative matrix and tensor factorizations. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.*, 92(3), 708-721. <https://doi.org/10.1587/transfun.E92.A.708>
- CoreNLP. A suite of core nlp tools (Stanford). <http://nlp.stanford.edu/software/corenlp.shtml> (accessed Dec 29, 2018)
- Cox, R.S. and Perry, K.M.E. (2011). Like a fish out of water: Reconsidering disaster recovery and the role of place and social capital in community disaster resilience. *Am. J. Community Psychol.*, 48(3-4), 395-411. <https://doi.org/10.1007/s10464-011-9427-0>
- Cutter, S.L. (2016). The landscape of disaster resilience indicators in the USA. *Nat. Hazards.*, 80(2), 741-758. <https://doi.org/10.1007/s11069-015-1993-2>
- Cutter, S.L., Emrich, C.T., Mitchell, J.T., Boruff, B.J., Gall, M., Schmidlein, M.C., Burton, C.G. and Melton, G. (2006). The long road home: Race, class, and recovery from Hurricane Katrina. *Environment: Sci. Policy Sustainable Dev.*, 48(2), 8-20. <https://doi.org/10.3200/ENVV.48.2.8-20>
- Darwin. (2015). Deliverable D1.1 consolidation of resilience concepts and practices for crisis management. <http://www.h2020darwin.eu/project-deliverables>
- Darwin. (2017). Deliverable D 2.1 Generic resilience management guidelines. <http://www.h2020darwin.eu/project-deliverables>
- Gordon, J., Aguilar, S., Sheng, E., and Burns, G. (2017). Structured generation of technical reading lists. *Proc. of the 12th Workshop on Innovative Use of NLP for Building Educational Applications, Copenhagen, Denmark*, 261-270. <https://doi.org/10.18653/v1/W17-5029>
- Hahn, A., Mohanty, S.D., and Manda, P. (2017). What’s Hot and What’s Not? - Exploring Trends in Bioinformatics Literature Using Topic Modeling and Keyword Analysis. *Proc. of the 13th International Symposium on Bioinformatics Research and Applications, Hawaii*, 279-290. [https://doi.org/10.1007/978-3-319-59575-7\\_25](https://doi.org/10.1007/978-3-319-59575-7_25)
- Herrera, I., Grotan, T.O., Woltjer, R., Nevhage, B., Nilsson, S., Trnka, J. and Jonson, C.O. (2017). Applying resilience concepts in crisis management and critical infrastructures-the DARWIN project. *In 26th Conference on European Safety and Reliability (ESREL)* pp. 2137-2144. <https://doi.org/10.1201/9781315374987-324>
- Holling, C.S. (1973). Resilience and stability of ecological systems. *Annu. Rev. Ecol. Syst.*, 4, 1-23. <https://doi.org/10.1146/annurev.es.04.110173.000245>
- Homeland Security: Homeland Security Presidential Directive 21. Homeland Security Digital Library. <https://www.hsdl.org/?view&did=480002> (accessed May 15, 2018)
- Hosseini, S., Barker, K., and Ramirez-Marquez, J.E. (2016). A review of definitions and measures of system resilience. *Reliab. Eng. Reliab.*



- Syst. Saf. 145, 47-61. <https://doi.org/10.1016/j.res.2015.08.006>
- Kitchenham, B. (2004). *Procedures for performing systematic reviews*. Keele, UK, Keele University, 33, 1-26.
- Lin, C.J. (2007). Projected gradient methods for nonnegative matrix factorization. *Neural Comput.*, 19(10), 2756-2779. <https://doi.org/10.1162/neco.2007.19.10.2756>
- Linkov, I., Bridges, T., Creutzig, F., Decker, J., Fox-Lent, C., Kröger, W., Lambert, J.H., Levermann, A., Montreuil, B., Nathwani, J. and Nyer, R. (2014). Changing the resilience paradigm. *Nat. Clim. Chang.*, 4(6), 407-409. <https://doi.org/10.1038/nclimate2227>
- Lizarralde, G., Chmutina, K., Boshier, L., and Dainty, A. (2015). Sustainability and resilience in the built environment: The challenges of establishing a turquoise agenda in the UK. *Sust. Cities Soc.*, 15, 96-104. <https://doi.org/10.1016/j.scs.2014.12.004>
- Magis, K. (2010). Community resilience: An indicator of social sustainability. *Soc. Nat. Resour.*, 23(5), 401-416. <https://doi.org/10.1080/08941920903305674>
- Marchese, D., Reynolds, E., Bates, M.E., Morgan, H., Clark, S.S., and Linkov, I. (2018). Resilience and sustainability: Similarities and differences in environmental management applications. *Sci. Total Environ.*, 613, 1275-1283. <https://doi.org/10.1016/j.scitotenv.2017.09.086>
- McManus, S., Seville, E., Brunsdon, D., and Vargo, J. (2007). Resilience management: a framework for assessing and improving the resilience of organisations, New Zealand, Resilient organisations research report. [https://ir.canterbury.ac.nz/bitstream/handle/10092/2810/12606763\\_Resilience%20Management%20Research%20Report%20ResOrgs%2007-01.pdf?sequence=1&isAllowed=y](https://ir.canterbury.ac.nz/bitstream/handle/10092/2810/12606763_Resilience%20Management%20Research%20Report%20ResOrgs%2007-01.pdf?sequence=1&isAllowed=y)
- Miller, F., Osbahr, H., Boyd, E., Thomalla, F., Bharwani, S., Ziervogel, G., Walker, B., Birkmann, J., Van der Leeuw, S., Rockström, J. and Hinkel, J. (2010). Resilience and vulnerability: complementary or conflicting concepts? *Ecol. Soc.*, 15(3), 11. <https://doi.org/10.5751/ES-03378-150311>
- Moretti, F. (2013). *Distant reading*. Verso Books.
- Myers, C.A., Slack, T., and Singelmann, J. (2008). Social vulnerability and migration in the wake of disaster: the case of Hurricanes Katrina and Rita. *Popul. Environ.*, 29(6), 271-291. <https://doi.org/10.1007/s11111-008-0072-y>
- Nadkarni, P.M., Ohno-Machado, L. and Chapman, W.W. (2011). Natural language processing: an introduction. *J. Am. Med. Inf. Assoc.*, 18(5), 544-551. <https://doi.org/10.1136/amiajnl-2011-000464>
- Norris, F.H., Stevens, S.P., Pfefferbaum, B., Wyche, K.F., and Pfefferbaum, R.L. (2008). Community resilience as a metaphor, theory, set of capacities, and strategy for disaster readiness. *Am. J. Community Psychol.*, 41(1-2), 127-150. <https://doi.org/10.1007/s10464-007-9156-6>
- Pennebaker, J.W., Boyd, R.L., Jordan, K., and Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*, Austin, Texas, University of Texas at Austin.
- Ronzano, F. and Saggion, H. (2015). Dr. inventor framework: Extracting structured information from scientific publications. *Proc. of the 18th International Conference on Discovery Science, Alberta, Canada*, 209-220. [https://doi.org/10.1007/978-3-319-24282-8\\_18](https://doi.org/10.1007/978-3-319-24282-8_18)
- SJR. SCImago Journal & Country Rank. <http://www.scimagojr.com> (accessed May 28, 2018)
- Seligman, M.E. and Csikszentmihalyi, M. (2014). *Positive psychology: An introduction*. Flow and the foundations of positive psychology, Springer, 279-298.
- Sengupta, N. (2006). Fragmented landholding, productivity, and resilience management. *Environ. Dev. Econ.*, 11(4), 507-532. <https://doi.org/10.1017/S1355770X0600307X>
- Shittu, E., Parker, G., and Mock, N. (2018). Improving communication resilience for effective disaster relief operations. *Environ. Syst. Decis.*, 38(3), 379-397. <https://doi.org/10.1007/s10669-018-9694-5>
- Smith, G.P., and Wenger, D. (2007). *Sustainable disaster recovery: Operationalizing an existing agenda*. Handbook of disaster research, Springer, 234-257.
- Smith, G., Martin, A., and Wenger, D.E. (2018). *Disaster Recovery in an Era of Climate Change: The Unrealized Promise of Institutional Resilience*. Handbook of Disaster Research, Springer, 595-619.
- Southwick, S.M., Bonanno, G.A., Masten, A.S., Panter-Brick, C., and Yehuda, R. (2014). Resilience definitions, theory, and challenges: interdisciplinary perspectives. *Eur. J. Psychotraumatol.*, 5, 25338. <https://doi.org/10.3402/ejpt.v5.25338>
- Steiner, A. and Atterton, J. (2014). The contribution of rural businesses to community resilience. *Local Econ.*, 29(3), 228-244. <https://doi.org/10.1177/0269094214528853>
- Sun, L. and Yin, Y. (2017). Discovering themes and trends in transportation research using topic modeling. *Transp. Res. Pt. C-Emerg. Technol.*, 77, 49-66. <https://doi.org/10.1016/j.trc.2017.01.013>
- Tausczik, Y.R. and Pennebaker, J.W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *J. Lang. Soc. Psychol.*, 29(1), 24-54. <https://doi.org/10.1177/0261927X09351676>
- United Nations. (2015). *Sendai framework for disaster risk reduction 2015-2030*. New York, NY: United Nations International Strategy for Disaster Reduction. [https://www.preventionweb.net/files/43291\\_sendaiframeworkfordren.pdf](https://www.preventionweb.net/files/43291_sendaiframeworkfordren.pdf)
- Walker, B., Carpenter, S., Anderies, J., Abel, N., Cumming, G., Janssen, M., Lebel, L., Norberg, J., Peterson, G.D. and Pritchard, R. (2002). Resilience management in social-ecological systems: a working hypothesis for a participatory approach. *Conserv. Ecol.*, 6(1), 14. <https://doi.org/10.5751/ES-00356-060114>
- Wilson, G.A. (2012). Community resilience, globalization, and transitional pathways of decision-making. *Geoforum*, 43(6), 1218-1231. <https://doi.org/10.1016/j.geoforum.2012.03.008>
- Wilson, G.A. (2017). 'Constructive tensions' in resilience research: critical reflections from a human geography perspective. *The Geogr. J.*, 184(1), 89-99. <https://doi.org/10.1111/geoj.12232>
- Zoraster, R.M. (2010). Vulnerable populations: Hurricane Katrina as a case study. *Prehosp. Disaster Med.*, 25(1), 74-78. <https://doi.org/10.1017/S1049023X00007718>