Using Semantic Drift on Social Media for Event Detection, Differentiation and Segmentation

by

Nataliya Tkachenko

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Computer Science

June 2019

Contents

Glossary

categorization Is the central issue in cognitive linguistics, which implies that concepts can be grouped into categories based on the common properties. 38

citizen science Is a process of public participation in scientific research, when the later is conducted, in whole or in part, by amateurs. 4

computational sociolinguistics Is a new approach to the study of language variation and change, which lies at the intersection of linguistics and computer science. 27

cosine similarity Is a measure of similarity between two vectors that measures the cosine of the angle between them. 57

distributional semantics Is a research area that studies theories and methods for quantifying and categorizing semantic similarities between linguistic items based on their distributional properties in large samples of language data (corpora). 40

folksonomy A user-driven approaches for classifying online content into categories with help of various metadata elements, such as tags, keywords and interest groups. xvii

gloss Is a series of semantic explanations, such as definitions or context examples. 45

HSO Hirst-St-Onge semantic relatedness measure, which classifies directional relations in the Wordnet hierarchy. 45

image schema Is a term from cognitive linguistics, corresponding to the recurring structures in our cognition which establish patterns of understanding and reasoning. They can result from bodily interactions and linguistic experiences. 30

JCN Jiang-Conrath semantic similarity measure, which subtracts the information content of the least common subsumer from the sum of the information content of the individual concepts being analysed. By default, the information content of concepts is derived from the sense-tagged corpus SemCor. 45

LCH Leacock-Chodorow semantic similarity index, measuring the shortest path between two concepts and scaling that value by the maximum path length in the 'is-a' hierarchy in which they occur. 45

LESK Banerjee-Pedersen semantic relatedness measure, which uses the text of a gloss as a unique representation for the unifying concept and subsequently assigns relatedness by finding and scoring overlaps between the glosses of the two concepts. 45

lexeme Is a basic lexical unit, consisting of one or several words, elements of which do not convey the meaning of the whole unit independently. 7

LIN Lin semantic similarity measure, which scales the information content of the least common subsumer by the sum of the information content of the individual concepts being analysed. By default, the information content of concepts is derived from the sense-tagged corpus SemCor. 45

metaphor Is a figure of speech used for rhetorical effects, which translates properties of one object or concept onto another thus referring to the hidden similarities between them. 31

metonym Is a figure of speech where a concept is substituted by its structural or functional part with which it is closely associated. 38

morpheme Is the smallest meaningful or functional unit in linguistic structures. 31

nowcasting Is the prediction of the phenomena in (near)present or very recent past time. 6

PATH Path-related semantic similarity index equal to the inverse of the shortest path length between two concepts. 45

polysemy Is a property of the lexical signs to possess multiple meanings. xvii

prototype Is a cognitive reference point for all representatives of the meaning of a word or of a category. 38

RES Resnik corpus-based similarity index of the concept specificity. 45

semantic drift (also: *semantic change* and *semantic instability*) is a longitudinal forms of change of lexemic meanings when old meaning is substituted by the new one(s). 9

shuffled correlation Is the result of the randomization approach one would expect to see for the sample values if there was not any correlation between the two variables at the population level. 52

soft sensors (also: *human sensing*) is the process during which humans can themselves act as sensors for data collection in a variety of scenarios, but the most frequently in the online contexts. 5

spatial autocorrelation Is the presence of systematic spatial variation in a variable due to the tendency for locations that are close together to have similar values. 54

synonymy Is a linguistic phenomenon where word or phrase means exactly or nearly the same as another word or phrase. 37

synset Is a grouping of synonymous words that express the same concept. 38

word2vec Is a group of related models that are used to produce word embeddings. 57

WUP Wu-Palmer semantic similarity index, measuring the path length to the root node from the least common subsumer of two concepts that is the most specific ancestor concept they share. This value is normally scaled by the sum of the path lengths from the individual concepts to the root. 45

Acronyms

EMDAT The International Disaster Database, CRED. 15

GloVe Global Vectors for word representations. 57

GNN Graph Neural Network. 49

ICT Information Communication Technology. 26

LIDAR light detection and ranging (remote sensing). 6

LISA local indicators of spatial autocorrelation. 92

LSA Latent Semantic Analysis. 57

MAUP modifiable areal unit problem. 78

PPMI Positive Pointwise Mutual Information. 57

SGNS skip-grams with negative sampling. 57

SMM social media monitoring. 24

UGC user generated content. 25

VGI Volunteered Geographic Information. 26

WSD word sense disambiguation. 57

WSI word sense induction. 57

List of Tables

List of Figures

Acknowledgments

This thesis is a product of a wonderful journey, and would not have existed without help and support of so many incredible people. First and foremost I would like to thank my supervisors, Professor Stephen Jarvis and Professor Rob Procter, who enabled my participation in this truly exciting CDT program of the Science of Cities. I am extremely grateful for their support, criticism, encouragements, ideas, suggestions, opportunities and lots of good memories, which filled the years of my PhD studies.

I would like to thank all the fellow students (especially the first two cohorts, with whom we have shared participation in so many events and activities) of the Warwick Institute for the Science of Cities for creating such a great team, full of creative and stimulating energy; Some special thanks go to Neha, Vikki, Adam G, Malkiat, Zakiyya, Jamie and Mo. An additional lot of massive thanks also goes to our amazing administrators Yvonne, Katy and Larry, who - let's be honest! - have always provided much more and beyond mere administrative support.

I would like to thank some external academics, who contributed to shaping my research project and were kind enough to invite me to their meetings, reading groups and internship programs, specifically Professor Kasia Jaszczolt from the University of Cambridge, who introduced me to the concept of 'truth-conditional semantics' and explained "what it means to be post-Gricean in age of the corpus linguistics" and Professor Dagmar Divjak from the University of Birmingham, who sparked my interest in something called 'meaning construal'. I am sending my thanks to the fellows of the Alan Turing Institute, specifically Barbara McGillivray, Scott Hale, Dong Nguyen, Suzy Moat and Tobias Preis, who invited me to take part in the various semantic drift and urban analytics related activities. I'd like to thank folks from the British Geological Survey, Environment Agency and Met Office, who have been incredibly supportive (and patient!) with all my ideas and questions. My special thanks go to Stephanie Bricker, who invited me to spend a week at the BGS with her and has always been there helping me to shape my project and encouraging to be inventive.

I would like to mention and send some special thanks to my immediate (but geographically distant) family in Ukraine (my dearest Grandma, Mom, Auntie, Cousin and his family) and to newly acquired (and geographically close) family in the UK (my lovely Parents-in-law, Brother-in-law and his family). Without your support and belief in me nothing would have happened.

Mega thanks to my beloved husband and partner in all the crimes I initiate. I am extremely lucky I have met you. You know that.

To Mia, my Belovest Baby Daughter.

Declarations

This thesis is the author's own work and has not been submitted for a degree at another University.

Some parts of this thesis have been published by the author, some others have been submitted and currently under review or in preparation:

1. Tkachenko, Nataliya, Bricker, Stephanie and Jarvis, Stephen A. (2015) To dig or not to dig? Place and perception in subsurface housing. Proceedings of the Institution of Civil Engineers - Engineering Sustainability

2. Tkachenko, Nataliya, Procter, Rob and Jarvis, Stephen A. (2016) From Kontrollverlust to Kontrollgewinn : rediscovering a meso-level normative hierarchy for integrated water management in England using on-line activities. In: Tenth International AAAI Conference on Web and Social Media, Cologne, 17-20 May 2016

3. Tkachenko N, Procter R, Jarvis S. 2016 Predicting the impact of urban flooding using open data. R. Soc. Open Sci.3: 160013

4. Tkachenko, Nataliya, Zubiaga, Arkaitz and Procter, Rob (2017) WISC at MediaEval 2017 : multimedia satellite task. In: MediaEval 2017 : Multimedia Benchmark Workshop, Dublin, Ireland, 13-15 Sep 2017. Published in: Proceedings of the MediaEval 2017 Workshop

5. Tkachenko N., Procter R., Jarvis S. (2017) How do eyewitness social media reports reflect socio-economic effects of natural hazards?. In: Ciampaglia G.,

Mashhadi A., Yasseri T. (eds) Social Informatics. SocInfo 2017. Lecture Notes in Computer Science, vol 10540. Springer, Cham

6. Tkachenko N, Jarvis S, Procter R (2017) Predicting floods with Flickr tags. PLoS ONE 12(2): e0172870

7. Tkachenko N, Procter R, Jarvis S. Immunobiological properties of the words we use [in preparation]

8. Tkachenko N, Procter R, Jarvis S. Using semantic drift to estimate spatial navigation strategies during flood events [in preparation]

9. Tkachenko, Nataliya, Chotvijit, Sarunkorn, Gupta, Neha, Bradley, Emma, Gilks, Charlotte, Guo, Weisi, Crosby, Henry James, Shore, Eliot, Thiarai, Malkiat, Procter, Rob and Jarvis, Stephen A. (2017) Google trends can improve surveillance of Type 2 diabetes. Scientific Reports, 7. 4993

10. Tkachenko N, Procter R, Jarvis S. Change Matters: Using Asynchronous Heterogeneous Data to Quantify Wellbeing [under review]

11. Tkachenko N, Procter R, Jarvis S. Detecting flooded locations with help of postdemographic traces [under review]

12. Tkachenko N, Procter R, Jarvis S. Using social media as an indicator of floodplain housing market moves [under review]

Abstract

With observable paradigm shift in computer science from *predictive* modeling to the *generative* one, it became important to maximise exploration of the pathways towards *useful* data production. With currently dominating statistical and compositional data augmentation strategies, opportunities also emerged for more application-driven routes. The main value of such approaches lies in their capacity to offer insights into context or event specific data productions, currently overlooked by more *topologically neutral* machine learning approaches. The purpose of this thesis is therefore to provide empirical evidence for useful data generation by dynamic event-specific lexical semantic resources.

Various Web 2.0 applications due to their popularity have been accumulating large amounts of semantically rich metadata, which became readily available and easily exploitable. Tags, usually consisting of a single word, are one type of such data. Tag uses can vary largely across systems and platforms; Also known under the term folksonomy, tags are usually non-hierarchical and open-ended, thus reflecting users' unique perspectives regarding various contexts, or *resources*. This platform-enabled liberty of expression, however, has led to situations of frequent *semantic ambiguity* due to spelling mistakes, morphological variations, polysemy, multilingualism or inaccurate tag-to-resource associations. As a consequence, tag spaces are often regarded as inconsistent, noisy and hardly reliable data sources.

Recent surge of interest amongst distributional semanticists in long- and short-term fluctuations of word meanings on social media has suggested routes for successful temporal sense disambiguation, thus inviting discussions around useful

real-world applications for such emerging data resources. One of such applications - event analytics from the crowd behaviour perspective - is gaining an increasing attention from researchers and practitioners, especially in the fields of operations and situational management. Pursuing pragmatic aims of event *detection*, *differentiation* and *segmentation*, this application domain is represented predominantly by repetitive catastrophic events (such as natural hazards), during which directly or indirectly exposed populations tend to share their situational experiences on social media.

This thesis consists of three main parts, each corresponding to specific problem in event analytics: (i) detection, (ii) differentiation and (iii) segmentation. In the first part I used the concept of ontological semantic proximity on the words-candidates for semantic drift in order to highlight the dynamics of their semantic oscillations within event-specific category (i.e., *flooding*). In my second experiment I followed on these initial findings and performed an analysis verifying whether semantically unstable lexical material can augment our knowledge about main sub-types of floods, such as 'slow' (e.g., groundwater and pluvial floods) and 'fast' (surface water and riverine floods) ones. In my third experiment I employed combined lexico-visual modalities of the crowdsourced material to reconstruct changing perceptions of flood events in order to understand how event severity can or cannot determine situationally resilient behaviours.

Chapter 1

Introduction

1.1 Background

"There is nothing permanent except change"

Heraclitus, c.535-475 BCE

Better data does make for better decisions. This axiom can be frequently encountered in various high-profile publications, covering various advances in modern data analytics, including big data statistics or artificial intelligence. And whilst in most cases it is given that 'big' means 'more', the question frequently remains how to increase the volumes of *useful* data, which would enable more granular, better fitted and overall more trustworthy insights into *life* of various *objects* or *citizens* of the physical or social worlds.

According to Stanford Encyclopedia of Philosophy [SEP, 2002], the variety of the world we are surrounded with is conditioned not only by the assortment of their ordinary *objects* or *citizens*, but are also greatly dependent on the "sort of things that happen to or are performed by them". In the late 90s this view turned into a focal debate across several disciplines, studying human languages, perception and action in conditions of various events - static or dynamic ones, mental or physical, as well as positive or negative ones.

In recognition of such complexity from the computational perspective, back in 2002 by the Stanford Professor David Luckham introduced complex event analytics as an exciting interdisciplinary arena of the data science, which has not merely provided organizations with new ways to analyze data patterns in real-time, but also opened numerous and diverse avenues for the academic community to reflect upon, and explore. And although event definition itself was set in motion by a wave of industrial projects back in the 1990s (predominantly in the areas of active databases and discrete event simulations) its formal introduction did not happen until 2002, specifically in the book "The Power of Events", which presented complex event analytics as a "[...] set of tools and techniques for analyzing [...] the complex series of interrelated events [...]" [Luckham, 2002, p.xvii] and not only defined generically its terminology and prototype execution models for the very first time, also it acknowledged the naissance of big data analytics, with its belief in the enhanced analytical power of multiple data sources to provide insights into the most complicated phenomena.

Throughout Luckham's book, various natural world events that can be regarded as 'complex' are being defined in their broadest sense, irrespective of the domain of origin or application. The only condition for their relevance to complex event analytics is the presence of vast heterogeneous amounts of information available to describe them, sometimes referred to as 'event clouds' [Luckham, 2002, p.28].

Event cloud information can originate from both structured (financial data feeds, hydrological and meteorological sensor networks, IoT or ground-based instrumentation) and semi- or unstructuted data sources (news items, text messages, social media posts, user-generated content for medical or traffic reports, etc.). From the data politics perspective, these streams can be also re-defined as *official* (i.e., *authoritative*) and *crowd-generated* ones. Combination of official measurements with volunteered observations is particularly common in fields where the failure of authoritative data sources can cause incorrect or imprecise predictions, resulting in significant socio-economic consequences. One such example is the domain of weather prediction.

As a consequence of the demand for complementary data streams (not least facilitated by the advent of Web 2.0), different authoritative organizations started implementing various data collection platforms that would enable the diversification of traditional, historically evolved scientific observations (which are well described in the book [PM, 2015]). One example of such an initiative is the Weather Observation Website (WOW) [MO, 2018], launched in 2011 by the Met Office Public Weather Service, supported by Royal Meteorological Society and the Department for Education, for weather observers across the UK. The purpose of the website was to provide a platform for the sharing of current weather observations from all around the globe, regardless of where they come from, level of detail or the frequency of reports. The observations can be collected using specially-designed digital, scientific or wireless weather stations or just by looking out of the window or sending in a photo. It is hoped that the website will encourage further growth in the UK's weather enthusiast observing community and help educate children about the weather and that this will become the UK's largest source of weather observations. These initiatives are usually moderated by a network of voluntary observers, who contribute daily climatological readings; During the month these records go through quality checking before being stored in the permanent archive and forming an important part of the climatological record for the UK. The Met Office has an interest in knowing where the weather is having a significant impact on people, infrastructure, transport and other activities, so WOW includes a range of variables, including temperature, rainfall rate, present weather, wind speed/direction, humidity, pressure (MSLP), snowfall, soil moisture and a number of the media-specific entries, such as continuous webcam observations, photos, magnetometers, etc.

A similar initiative was launched around the same time at the British Geological Survey and was explicitly defined as citizen science, combining the entire cluster of volunteered activities. The overall term is used for projects in which individuals or networks of volunteers - many without specific scientific training - perform or manage research-related tasks such as observation, measurement or computation. The use of such citizen science networks therefore allows scientists to accomplish research objectives more feasibly than would otherwise be possible; and, in addition, these projects aim to promote public engagement with research, as well as science in general. Examples of citizen science initiatives include crowd maps of GeoExposures [GE, 2011] to geological hazards, such as landslides or flooding, which are reported in a simple form of location and evidence and which can be supplied in the form of images, web or video links, as well as simple text descriptions. Other applications are more hazard- or domain specific, for example, [MS, 2011], iGeology [IG, 2011] platforms or landslide [RL, 2011] or earthquake [EI, 2011] reports. Some activities are purely analog, and encourage people to collect samples of volcanic ash in order to contribute to the collection of evidence of the distribution of ash falls [AS, 2011] or to share their subjective experiences during an earthquake via short questionnaires [SS, 2011].

The main concept behind all such initiatives is the idea of active citizen sensing, which assumes active or voluntary design, participation or contribution to the task, often linked to the concept of so-called *proactive* behaviours, which most often assume personal interests or concerns (for example, of being flooded or protecting investment against being damaged by geological hazards, such as fluctuating groundwater levels), and which have been defined as ‘psychologically comforting’ and ‘healthy’ behaviours [Rotenberg, 2009].

However, apart from more active forms of data collection, modern media offers also entire arrays of data streams, which can be useful *potentially* as they result from mediated human activities not designed for the purposeful data production. The idea of using social media as a data source stemmed from the fact that social media channels, such as Facebook, Twitter, YouTube and Flickr started being increasingly used to promote situational awareness as an increasing number of people look to social media as an additional, more immediate and readily available source of information from the late 2000s. Since then, the growing importance of social media and its growing data repositories started being recognized by authoritative institutions, which prompted them to explore the full potential offered by such alternative data sources and resulted in the design of such platforms as WOW (MetOffice) and GeoSocial [GS, 2011] (BGS).

The power of these prototype platforms at the moment lies primarily in their interactivity, as, for example, GeoSocial allows users, including BGS scientists, to visualize geoscience-related posts from social media sites, such as Twitter, in real time, where social media provide a different channel for gathering potentially useful information for scientific analysis from the public. Hence, their official statement on the website (as per August 2018) confirms that "the aim of GeoSocial is to explore whether BGS can make use of the wealth of information that is publicly available through such sites to help advance scientific understanding and provide better more timely advice."

The method behind GeoSocial data collection identifies potentially relevant postings about different geoscience themes (such as landslides, flooding, volcanic eruptions and earthquakes) from Twitter and, if they can be located, displays them on a map interface. The application currently relies on keyword filtering and only searches for English terms, therefore it has been acknowledged that some relevant postings may not be retrieved where they do not meet the predefined search criteria. However, even with the filtering levels, the posts that are displayed in real time are still essentially the content provided by the Twitter platform and hence are beyond the BGS's quality checking procedures, so may contain offensive material or other unconfirmed information (such as lies or rumours).

The emergence of such tools and platforms signified the acknowledgement of novel and alternative data sources by authoritative and government institutions around the world, even though they are still unsure of how useful they can be. Thus, BGS defines that one of the routes for developing understanding could be via novel data analytical techniques, such as machine learning, exploration of novel data collection instruments or crowd verification. Used on new data sources and with help of the novel analytical techniques, these approaches would enable understanding of the usefulness of these types of information by [hazard] scientists and experts, when used either independently or alongside structured official data holdings, field observations and models.

1.2 Motivation

The main motivation behind this project is to understand the ways soft sensors (also known as human, or citizen, sensors) could enrich existing complex event analytics routines with semantic information contained in user-generated posts and entries (text, photographs), and answer specific outstanding questions that have been found challenging by more traditional data and modeling approaches.

In the scope of this work I therefore aimed to look into how unstructured data sources derived from social media can be rendered useful for geoscientists. I turned my attention to the sub-field of the natural hazards [floods in particular] as a prominent case of 'human-environment' interactions because of their firm positioning between *physical* and *behavioural* geographic events [Bunting and Guelke, 1979; Gold, 1980].

The main methods underlying contemporary, authoritative flood risk communication systems are computational, using interpolated temporal measurements of hydrological sensor networks (precipitation, surface water, groundwater and tidal gauges), combined with the spatial topographical floodplain designations, which are based on historic geodetic measurements and more recent light detection and ranging (remote sensing) (LIDAR) scans of the earth surface or aerial photographs of flooded areas. The high cost behind data collection and processing, time delays required to get information extracted and, most importantly, lack of immediate interpretable mechanisms of real social impacts have rendered this techno-centric approach only partially efficient [Alexander, 1991]. Therefore the research community has searched for cheaper and faster alternatives, like citizen science, which have been in use to gather important ecological and environmental information since the beginning of the 20th century [Sachdeva and McCaffrey, 2018; Silvertown, 2009].

On the other hand, since the advent of the Web 2.0, social media is increasingly used for nowcasting and prediction of events in the social world [Moat et al., 2014], and some initial studies have also demonstrated its potential for monitoring and predicting phenomena in the physical world as well, such as storms, tsunamis, hurricanes and droughts [Preis et al., 2013; Peary et al., 2012; Freberg et al., 2013; Tang et al., 2015]. As observed, the advantages of using these data signals are due predominantly to their complementary nature. Thus, some studies reported on its potential strengths in areas where gauges or sensors are absent altogether [Aggarwal and Abdelzaher, 2012; Javelle et al., 2014], while others demonstrate its power when being used alongside existing sensor readings [Restrepo-Estrada et al., 2018], thus improving on the accuracy and precision of the resulting signal and, as a consequence, of the final warning message.

As event forecasting and monitoring is being increasingly characterized as a 'big data' problem, so the problem of data quality becomes apparent, where in the conditions of apparent data abundance, the *useful* signal that can be extracted, is either too sparse, or geographically inconsistent. For instance, while Portuguese-speaking Twitter can generate vast amounts of georeferenced signals, other Twitter communities can boast far less useful information (1-5 per cent) due to privacy concerns and social media participation preference trends [Brogueira et al., 2016].

The linguistic component of social media messages (i.e., text, tags, annotations, etc.) is usually the primary filtering mechanism that enables data selection for advanced event modelling, as it usually contains a direct description of the event. These extracted data signals are further validated by two other important metadata elements, *timestamps* and *geolocation*, which define the final usability of the extracted data by helping to verify whether messages have been sent from the nearby area or around the same time the event broke out. Very often, whilst event data streams on social media seem to be generating huge volumes of data, the amount of actually *useful* post-filtered information can be significantly reduced, often due to the small fraction of geolocated media on the social web. It therefore becomes important to understand the mechanisms that could help to increase the volumes of useful information that can be extracted for event modelling purposes.

One such method could be expansion of initial linguistic filtering by extending the volume of linguistic descriptors that could be potentially related to the event and are either represented in a more significant fraction of the data uploaded to the platform or are known to have a higher fraction of georeferenced data, for example, originating from trusted profiles that have location services enabled on their devices. For this reason, I decided to turn my attention to the concept of *semantic drift* (also known as semantic change, or shift, in the literature on computational distributional semantics); The main principle behind this approach is that some lexemes possess several potential meanings and can be used as event descriptors themselves, thus increasing the volume of useful risk-related information, improving the event sensing potential if posted prior to the direct event-describing lexemes and reflecting behavioural patterns around official flood risk communication.

As the main case study, the *problem of flooding* was therefore selected due to its properties as a socio-natural phenomenon, which can substantially benefit from the formalized complex event analytics approaches:

1. The *uncertain nature* of the phenomenon. In the UK flooding is considered to be one of the biggest problems that the country is facing today, with climate projections suggesting that increase in total rainfall will provoke major, more frequent and less predictable flood events [Wilkinson et al., 2015];

2. It has been widely acknowledged that the issue of flooding as a typical example of natural hazards has been predominantly approached from its geophysical, climatological and meteorological perspectives, and far less from its social dimension, which poses a problem as "natural hazards cannot be considered

independently of the individuals and groups that they afflict" [Gold, 1980, p.203]. This research gap was identified in the early 1980s and has not been successfully addressed yet [Mackay et al., 2015], primarily due to the lack of data as compared with information on the physical extent and magnitude, data on human perception of, and response to, flooding has been considered to be much harder to find [Beven, 2007];

3. Complex nature of the different *types of the phenomenon*, lack of knowledge about which affects both accurate predictions and efficiency of selected measures [Wilkinson et al., 2014]. Whilst the most known flooding type is when river waters overtop their banks, known as *fluvial*, or surface water flooding, there are also exist less known and less well understood cases of *pluvial* and *groundwater* flood types. Pluvial or 'precipitation-related' floods usually occur following short intense downpours that cannot be quickly enough infiltrated into the ground or evacuated by drainage systems. Groundwater flooding occurs as a consequence of elevated levels of the groundwater table, which may happen days or even weeks after heavy or prolonged rainfalls, causing water stay in streets or inside properties for a long time, sometimes up to several months. *Sewage floods* can occur as a result of both surface and groundwater overflows, usually as a result of pipe misconnections or blockages. Diversity of the sampling mechanisms behind various types of flooding prompts researchers to turn their attention to non-orthodox data sources, capable of capturing the tacit aspects of these distinct flood typologies;

4. Information on detailed human behaviour during floods is still scarce [Aerts et al., 2018]; It has been already mentioned by several researchers that novel flood warning mechanisms could substantially benefit from the data, for example, on human mobility as an indicator of risk perception during various stages of the hazard [Wang et al., 2016]. As traditional data collection tools cannot stand up to the challenge [ben, 2009], there is therefore a scope for alternative data sources, containing behavioural signals, to be tested for their usefulness alongside more traditional, physical instruments [Baldassarre et al., 2013; Cologna et al., 2017].

The complexity of processes behind the evolution of meteorological and hydrogeological events into flooding hazards and associated difficulties with advance risk estimations also propagates uncertainties among actors and beneficiaries within the system of environmental regulation, such as property developers [har, 2016] and insurers [Petrolia et al., 2013], which could potentially benefit from new methods able to capture complex human experiences of the risky situations, collective memory and perceptions [Birkholz et al., 2014]. This hypothesis also resonates with the recent statement made by the Environment Agency (2015) [TNA, 2015], acknowledging the limitations of current, predominantly technocentric, flood risk management approaches based on the traditional data sources and methods: "Recorded flood outlines contain the individual records of historic flooding. These records show flooding to the land and do not necessarily indicate that properties were flooded internally. Absence of a flood event does not mean that the area has never flooded, only that we do not currently have records of flooding in this area [...]".

1.3 Research questions

The main principle of complex event analytics is to get better information to enrich situational analytics and respond to emerging hazards as quickly as possible [Etzion and Niblett, 2011], with flood monitoring on social media as the chosen case study. Following abovementioned problems in flood risk analytics (1-4), I have identified the following research questions for in-depth investigation:

1. *Can we detect events (e.g., flooding) with help of semantic drift on social media?*
2. *Can semantic drift on social media help to differentiate types of flood events?*
3. *Can alternative tags (i.e., the ones, which are not direct event descriptors) help us to distinguish different stages of flood events?*

1.4 Thesis contributions

This thesis makes the following specific contributions:

1. Proposes three alternative methods for behavioural event analysis as compared to more traditional authoritative data sources used in flood risk monitoring and management;
2. Critically evaluates the *complementarity* between authoritative sensor networks and soft (social) sensors;
3. Introduces practical applications for the popular in data science concept of semantic drift (also known as *semantic instability* in computational linguistics) by adapting it to the requirements of *situational analytics.* I propose three

types of such pragmatic applications: (i) *transient* semantic drift, (ii) *spatial* semantic drift and (iii) *cognitive* semantic drift, each well adapted to answer specific question in flood risk analytics.

1.5 Thesis structure

The remainder of this thesis is structured as follows:

Chapter 2 (Literature review): In this Chapter I present a detailed overview of the literature sources covering the principal research trends in the field of human perception of natural hazards and how social media data could step in in order to address outstanding event-focused questions in the example of flooding phenomena. I start off with the main approach to the classification of the natural hazards, demonstrating diverging trends between geophysical and the social systems. In order to better understand currently underrepresented research field of the natural hazard analytics from a human perspective, I look at broader theories, arising from experiences of the surrounding environment, and how have they evolved from being purely reflexive towards becoming computationally orientated (AI and natural scene statistics). Following this, I attempt to position natural hazard analytics in the context of complex event analytics and continue by uncovering contemporary trends in the latter domain and how it models hybrid socio-technical systems. After bringing into focus social media as a data source, I describe the current state of art such 'social sensors' play in natural hazard analytics and present outstanding research avenues for the computational behavioural component of natural phenomena.

Chapter 3 (Methodology): In the scope of this Chapter I present the main arguments behind the choice of the research questions selected for further empirical verification. I start off by presenting semantic drift as a main analytical tool for analysis of the human perception of flooding on the multimodal platform Yahoo! Flickr. This discourse is followed by hypotheses theoretically based on state-of-the-art approaches in the field of distributional semantics and is repurposed for further practical verification in context of the flood events. The second half of the Chapter is dedicated to the theoretical reasoning behind selection of the *primary* (e.g., variables-candidates for semantic drift) and *secondary* (authoritative flood monitoring datasets) data sources for the three research questions of this thesis.

Chapter 4 (Event detection with semantic microchange on social media) is based on the assumption that user generated content (UGC) in social media postings and their associated metadata such as time and location stamps can be useful for providing valuable operational information during natural hazard events. The implications of the latter are such that, in a purely additive sense, they can provide much denser geographical coverage of the hazard as compared to traditional sensor networks, whilst also being able to provide what physical sensors are not able to do; Notably, by documenting personal observations and experiences, they directly record the impact of a hazard on the human environment. As choices of semantic tags in the current methods are usually reduced to the exact name or type of the event (e.g., tags 'Sandy' or 'flooding'), the main limitation of such approaches remains their mere *nowcasting* capacity. In this analysis I therefore make use of polysemous tags of images posted during several recent flood events to demonstrate how such volunteered geographic data signals can be used to detect event before the direct event descriptors.

Chapter 5 (Event differentiation with spatial semantic drift) covers the already known and well reported problem of noise in social media data sources, where words and expressions [can] often mean something different not directly related to the event in question. In addition to this, there is also a problem of data precision, which often prevents us from finding out exact types or subtypes of events we want to get more detailed insights into. Nevertheless, current research efforts in this area of analytics fail to appreciate an outstanding opportunity to explore how such *noise* can be exploited for gaining deeper insights into complex flooding phenomena, where traditional filtering by the keyword 'flood' is hardly helpful. To address this gap, in the scope of analysis presented here I use the concept of *semantic drift* in order to understand how ontologically related words on multimodal social media platforms can generate additional pools of useful, but currently underexploited data components. My results demonstrate that alternative keywords are able to differentiate two main, according to the USGS classification system, types, specifically *riverine* (river flow and surface water floods in our analysis) and *flash* (precipitation and groundwater) floods.

Chapter 6 (Exploring potential of semantic drift for event segmentation). As *semantic drift* is a known research category of distributional semantics for its capacity to demonstrate gradual long-term changes in meanings and sentiments of words, its empirical performance is nevertheless largely determined by the corpus composition. In Chapter 4, which used ontological relationships between words and phrases, I have already established that there also exist certain types of semantic *microchanges* on social media, emerging around natural hazard events, such as floods. My previous results confirmed that such alternative lexical material can be used to detect floods before their outbreak and to increase the volume of 'useful' georeferenced data for event monitoring. In this final experimental Chapter I use deep learning in order to determine whether pictures associated with ‘semantically drifted’ social media tags reflect changes in the event severity or are a reflection of the people’s reaction to the authoritative flood risk communication. The results show that alternative tags do follow the pattern of the direct event descriptors and are indeed sensitive to the evolving severity of the hazard. They also have more complex composition, ranging from more focused to less focused scenes, which can be used as statistical indicators of flood risk severity.

Chapter 7 (Results and Discussion) summarises findings from the previous chapters and critically reflects on emerged theoretical questions and outlines future directions for some similar and alternative practical implementation(s).

Chapter 8 (Conclusions) presents general overview of results, what implications they may have for the field of natural hazards analytics in context of the growing interest towards socio-environmental studies.

Chapter 2

Literature review

Synthesis

In this Chapter I present a detailed overview of the literature sources covering the principal research trends in the field of human perception of natural hazards and how social media data could step in in order to address outstanding event-focused questions in the example of flooding phenomena. I start off with the main approach to the classification of the natural hazards, demonstrating diverging trends between geophysical and the social systems. In order to better understand currently underrepresented research field of the natural hazard analytics from a human perspective, I look at broader theories, arising from experiences of the surrounding environment, and how have they evolved from being purely reflexive towards becoming computationally orientated (AI and natural scene statistics). Following this, I attempt to position natural hazard analytics in the context of complex event analytics and continue by uncovering contemporary trends in the latter domain and how it models hybrid socio-technical systems. After bringing into focus social media as a data source, I describe the current state of art such 'social sensors' play in natural hazard analytics and present outstanding research avenues for the computational behavioural component of natural phenomena.

2.1 Natural hazard perceptions

2.1.1 Approaches to the natural hazard classification

Traditionally, the term 'natural disaster' has been remarkably difficult to define and, as a consequence, there exist several approaches for classification.

According to the broader classification, natural hazards fall into the category of *negative* events, which according to [Iliev et al., 2016], are "less numerous, but more diverse when compared to positive events, which are much more common."

According to [Burton and Kates, 1964], natural disasters can include various environmental phenomena, including blizzards, floods, tornadoes, earthquakes, volcanic eruptions, fungal diseases, infestations, etc. From the perspective of physical geographers primarily interested in climatological and geological events as opposed to biological phenomena, a natural hazard is "an extreme geophysical event greatly exceeding normal human expectations in terms of magnitude or frequency and causing major human hardship with significant material damage to people and their possessions and possible loss of life." [Oliver, 1975]. However, a natural hazard would not have been a hazard if it existed independently from human settlements, the individuals they afflict and the locations of their activity, so according to [White, 1974, p.3] "[...]no natural hazard exists apart from human adjustment to it. It always involves human initiative and choice. Floods would not be hazards were people not tempting to occupy floodplains; By their occupancy people establish the damage potential and may well change the flood regimen itself."

From a purely geoscientific perspective, natural hazards are divided into groups according to the main mechanisms behind their emergence and evolution. According to The International Disaster Database, CRED (EMDAT) [IDD, 2011], the most common division, adopted in many countries and according to which national authoritative risk communication organisations are structured: *meteorological* (e.g., different types of storms, blizzards, heat waves, tornadoes, etc.), *hydrological* (e.g., different types of flooding, tsunamis and related multi-hazards) and *geological* (e.g., earthquakes, avalanches, landslides, volcanic eruptions, etc.) natural hazards.

The peculiar fact about natural hazards is that their boundaries are often blurred as they may emerge due to various interacting geophysical factors. This can lead to cases of multi-hazards or 'cascading hazards', such as, for example, flooding and landslides occurring after rain over wildfires [AghaKouchak et al., 2018]. Also, there are often cases of 'nested hazards', where one type of disaster may consist of several interrelated hazards of similar types but of different origins [Sophocleous, 2002]: If we look at flooding specifically as a natural phenomenon it is extremely complex and subject to change. Incidents are no longer restricted to obvious areas where a river or stream exists; for example, many urban floods are simply caused by huge amounts of rain falling very quickly (*flash* floods) in an area where the drainage system is unable to cope or due to unexpected underground basin recharge and rise of groundwater levels [Handmer and Proudley, 2007]. As a consequence, there is an ongoing motivation to understand how accurate our knowledge can be about natural hazard risk — its location, timing and duration — which is being fuelled by both the diverse and changing nature of natural hazards, but also by new, emerging datasets, methods and computational tools [Gaitan et al., 2016], which can be potentially useful for answering such longstanding questions.

2.1.2 Human responses to natural hazards

The problems posed by people's understanding of and response to natural hazards have long been of concern to both academic and policy-making communities. Geographical research into natural hazards has evolved steadily over time within a unified paradigm, traditionally process- and instrument-orientated [Brown and Damery, 2002; Blair and Buytaert, 2016]. This has served to give natural hazards research the considerable advantages of coherence and integration that contrasts markedly with other areas of behavioral geography. According to some authors, these benefits, nevertheless, have been largely "counterbalanced by a parochial outlook and narrowness of analysis" [Gold, 1980, p.202][Blair and Buytaert, 2016], and they claim that discussion of this matter must be postponed until the full, or at least more extensive, range of human responses to natural hazards is examined using new and emerging data sources, thus embracing more integrative, socio-technical mechanisms of natural hazards analytics.

Studies of how the natural environment defines human behaviour has shown a rapid development in recent decades, coinciding with the rise of the *behaviorist* tradition across several fields of the social sciences, including psychology, sociology, geography, anthropology, which, in turn, have also sparked an interest in more applied environmental design fields, such as architecture, urban and regional planning, and interior design. Fairly recently, this interest has also found manifestation in more computational disciplines, such as data science and information engineering, aiming to incorporate human agents' behavioural signals into complex models of various phenomena of the physical world [Filatova, 2015; Dawson et al., 2011].

Since complete ignorance of the existence of hazards is rare, some generally recurrent features can be noticed in studies of responses to natural hazards, which are the most often approached from either *evolutionary-cultural* or *personal* perspectives. The ways in which humans used to pass down stories through the ages helped cultures to cope when disaster struck and provided rich research information grounds for anthropologists and social scientists. In 1968, Dorothy Vitaliano, a geologist at Indiana University, pioneered the study of cultural myths that told of real geological events [Hamacher and Norris, 2010]. According to [Janif et al., 2016], this story represented a unique geological record of ancient eruptions. Similarly, McAdoo [McAdoo and Paravisini-Gebert, 2011] studied earthquakes that triggered tsunamis, looking at the social and cultural factors that made some geological disasters deadlier than others. Other scientists were using a similar strategy to study seismic events in the Pacific Northwest. [Atwater et al., 2005] was tasked with mapping earthquake risks across Northern California, Oregon and Washington, and was searching for information about previous earthquakes in those areas. However, written records dated back only about 200 years. With help of Japanese seismologists, he was able to link traditional native stories to historic records and start a productive collaboration. After the tsunami 2004 in Indonesia, during which hundreds have been killed or left homeless, event stories-instructions to run immediately to high ground if water recedes after an earthquake permitted residents of remote villages to survive with relatively few casualties, and as they gained in popularity among people, their geological merit began to grow as well. Since then, there have been more constructive dialogues between social scientists, natural scientists and engineers, which have led to more insights on how and why these disasters happened. Fairly recently, geologists have also begun to pay more attention to how indigenous peoples understood and prepared for disaster. It started to be recognized that these stories could ultimately help scientists prepare for cataclysms to come, also confirmed by the tsunami outcomes on the Andaman and Nicobar Islands, during which islanders who had heard the stories about the *Laboon tsunami* or similar mythological figures survived the tsunami almost unaffected, whilst many residents in the city of Port Blair, being outsiders with no indigenous tsunami warning system to guide them to the higher ground, became the main victims.

The second approach, covering personal experience factors was described by [Ittelson, 1976], who put forward four reasons why human reaction varies in this way:

(1) *Previous experience*: as natural disasters are not part of everyday lives, the way people perceive dangerous situations is conditioned by how long ago the situation took place, as rare events usually require re-adjustment from scratch because memories about them tend to fade away over time.

(2) *The role of personality*, about which little is yet known. It has been

hypothesized that people living in hazard zones have personalities similar to gamblers, willing to take risk of losses caused by natural hazards against the prospects of untroubled living.

(3) *Attachment to place*, which causes general unwillingness by the public to recognize negative implications. Events often require major readjustments in the ways people live, which they are usually reluctant to make. According to ecological concepts of the natural environment, human systems are seen as having their place within the natural order, where each individual needs to protect themselves, recognizing that the natural environment is unpredictable and that risk is always somewhere there.

(4) *Attitudes towards nature.* Information from the hazard environment is often ambiguous and, as a result, judgements are less accurate than is normally the case with other environments. For centuries, nature has been seen as beyond human control and, as a consequence, associated with divine will. In numerous theological works, fire, earthquakes, lightnings and floods have been seen as supernatural forces, that are used "to cleanse evil or as tokens of divine displeasure."

The latter reason is based on the known fact that people tend to associate their everyday experiences with symbolic values, hence their emotional investment into the landscapes they see and visit is no different. Such symbols can yield varying and sometimes conflicting emotions, such as pleasure, pain, melancholy or nostalgia. As a consequence, people's relations with the surrounding environment can also manifest itself in a myriad ways, where, in some instances, the emphasis is made on landscape preference, attraction and friendliness and, in others, it has been based on it being seen as a continuous source of risks and therefore provoking stress, avoidance or adjustment.

The aesthetic conflict towards various landscape experiences has been expressed in Auden's theory of *topoi* (or commonplaces), where pleasurable landscape experiences are neatly summarized under the terms *topophilia*, which is a primary category and has a more embracing meaning of a response of open and suggestive emotion towards particular landscapes, including those that have never been visited or that perhaps do not even exist. In this respect, according to [Betjeman, 1947] topophilia has powerful links to utopia and heaven, where the landscape stands for a degree of perfection, a person's preferred location, which by the act of association has been to a certain degree humanised.

Topophilia can be deeply personal, however, it can also be a shared experience; in addition, it can present images of a less positive kind, such as pain, suffering, fear, or desolation. And although such *topophobic* images might be less expressive or powerful than those ones of topophilia, they can be surprisingly consistent over time. For example, [Gold, 1980, p.118] describes that "...mountain, forest, river have only recently become landscapes of attraction, loved today for solitude which provides a counterpoint to the pressures of the modern human landscape. In the past adjectives such as 'awful', 'horrible', and 'hideous' were applied to the landscapes of nature upon which human art was unable to act."

In his analysis of Auden's theory, [Gold, 1980, p.118] also points out that there has been an interesting reversal in topophobic landscapes since industrialisation: as nature has become much better controlled, areas of wilderness, which previously has been viewed as dangerous shifted from being a word of topophobic content to one of topophilia, to the extent that the search for original naturalness has become a search for utopia in the minds of many [Lowenthal, 1964]. At the same time anthropogenic landscapes, particularly industrial ones, have turned into a topophobic concept. For example, the image of the dark, smoky and polluted industrial landscape (e.g., Blake's "dark Satanic mills..." (1804) became and persisted for decades as a dominating theme in attitudes towards cities, but at the same time it is the image that has stimulated contemporary concerns with conservation and landscape preservation. Regarding the examples of European topophilia [Glacken, 1967] demonstrated how contradictions between accepted landscape theories and scientific, geographical and economic development have led to changes in environmental attitudes. More recent studies have also attempted to demonstrate the role of antropogenically altered environment in enhancing unethical antisocial behaviours and crime [Lu et al., 2018].

Another theory was put forward by [Appleton, 1975], who combined the animalist arguments of Dewey with associationalist thinking to produce *habitat theory*, where he suggested that "in order to achieve evolutionary survival, man had to learn how to see without being seen." (p.239) This theory produced a framework of landscape symbolism that was organised around three main concepts of 'prospect', 'refuge' and 'hazard'. Because hiding was seen as absolutely crucial landscape behaviour for primitive societies, it was important to use surroundings in a way that allowed warning of an approaching animate or inanimate hazard. Therefore, according to the theory, if it were not for the existence of *hazards*, *prospects* and *refuges* would have no meaning; Thus, the concept of a hazard was turned into a cornerstone concept of the symbolism framework.

As well as attributing landscape with artistic meanings and connotations, symbols nevertheless are also products of the human mind that facilitate the communication of ideas and images about landscape, which are also firmly grounded in the culture in which they are found. [Panofsky, 1970] has claimed that symbolic meaning operates at three levels, which broadly parallel the three elements of landscape (actual physical features, associated activities, performed on the scene, and proper meanings or symbols). At a primary level Panofsky identified *factual meaning*, as when the representation of a particular natural feature is associated with the real landscape element and *expressional meaning*, in which certain psychological states are identified by their representation, for example, a face expression or a word. The secondary level was *conventional meaning*, which is approached through identification of symbols. For example, linguistic metaphor of 'standing water' can be used to describe *flooding*. However, in order to achieve that recognition, we require some knowledge of the traditions or habits of representation of that particular natural phenomenon in that particular geographical area. At the third level, the *proper symbolic level*, iconology is important for the interpretation of the landscape metaphor, specifically how meaning was formed in those particular settings. From this perspective, in order to comprehend the symbolic landscape, we must understand the basic values and attitudes of those who produced that landscape and those who hold images of it. In this respect language as a system of symbols is a particularly representative iconology as it is used for meaningful communication. However, this may consist of various sounds, gestures, or written characters that may represent objects, actions, events that can be either directly translatable or metaphorically encoded.

Environmental meaning creation has been also a concern of Gibson's *ecological psychology*, which suggested that the affordances of the environment, what it can offer to the animal, either good or ill, define its perception and all the subsequent activities [Gibson, 1979]. In its broadest sense, ecological psychology complements the traditional focus of environmental ecology, which is primarily concerned with energy transactions between the living systems and their environments by emphasising yet another angle of information transactions. In later works, Gibson used the term 'ecological' psychology to emphasise this animal-environment reciprocity for the study of problems of environmental perception, since he believed that analysing the surroundings was just as much a part of the psychologist's task as analysing the *perceivers* themselves, and hence that the physical concepts of the landscape and the biological concepts of the organisms would have to be tailored to one another in a larger system of mutual constraint.

In his complementary on Gibsonian theories of environmental perception, Nicolai Bernstein also presented an *ecological* approach to the problems of situated coordination and movement [Bernstein, 1967]. He pointed out that action cannot be studied without reference to the environment and that physical and biological concepts must be regarded together as a reciprocal system. This coupling of Gibson's ideas with those of Bernstein therefore have provided a natural basis for looking at traditional psychological topics of perceiving, learning and acting as activities of crowds rather than of isolated individuals.

In order to consider some broader influences on processes behind meaning attribution to geographical spaces, it is worth looking at theories of *environmental learning*, which are often in opposition to more instinct-based theories of ecological psychology. This school of authors argue that whilst there are multiple factors in play that can define human attitudes to the natural environment, landscape aesthetics perception is mostly defined by whether the person is *a visitor* or *a resident* (i.e., the *temporality* of presence) in an area as the desired landscape and the hazard zone are often one and the same thing [MacEachren, 1992]. They looked at opportunities to extend micro-scale studies of developmental psychologists, whilst also considering the influence of various socio-cultural factors on an individual's view of the natural and urban environment [Boulos et al., 2011; Dubey, 2016; Ruiz-Correa et al., 2017].

Theories based on the principle of the *temporality of presence* maintained that differences in the mental or behavioural characteristics of individuals are due primarily to environmental experience, which, in turn, resulted in several perceptual laws first proposed by the Gestalt psychologists [CHE, 2019]. Since Gestaltists believed that the brain had innate self-organising tendencies, they tried to explain [environmental] perception and cognition specifically from this perspective. They believed that the first stage in perception was the ability to distinguish the *foreground* (i.e., central) figure from its *background* (or simply 'ground' as they referred to it) and conducted a series of experiments that put forward the principles by which they believed such pattern recognition takes place: 'proximity', 'similarity' and of 'good continuation'.

According to the *proximity principle* of perceptual grouping, elements that are close together are perceived as a *group. Similarity* suggests that visually similar objects tend to serve as a basis for space fragmentation for the observer, whilst the *good continuation* principle allows for the primary perception of simpler shapes before they start constituting perceptually more complex formations. Although relatively well-structured, these principles did not find an empirical confirmation in subsequent works by [Elkind and Scott, 1962] and [Bower, 1965] on age differences between observers, and by [Schwitzgebel, 1962] and [Vernon, 1971] looking at educational and cultural differences. However, the idea that human ability to perceive objects or shapes as holistic figures against the background, reflecting that we learn to allocate them to a specific category of objects or shapes with which we are familiar, has recently been followed up by the work of Berkeley researchers on *natural scenes prediction* with help of computer-assisted classification algorithms [Stansbury et al., 2013]. Here, the authors pursue the idea that machine learning algorithms can classify natural scene objects into visual semantic categories just as well as the human visual cortex, which has been the subject of quite a substantial body of research in natural scene statistics. Starting off with the ability to quantify natural objects and shapes the field has evolved into an interdisciplinary area capable of negotiating with and confirming some of the most fundamental research hypotheses in psychology, specifically about the role of attentional mechanisms in natural scene perception.

According to recent findings, perception of scenes in the natural environment to a great degree is dependent on the awareness and consciousness (i.e., *presence*) of the observer, which are only possible in conditions of *attention* as natural-scene perception is not a pre-attentive process [Cohen et al., 2011]. Also, following this quantitative tradition, what happens is that humans tend to categorize scenes they encounter in their everyday life, such as park, workplace or the beach and these categories are assembled according to knowledge of the way objects co-occur in known (or *learnt*) natural scenes, and can also be used to distinguish one category of natural scene from another. Moreover, the results have confirmed that such inferences can also be made in the opposite direction, i.e., the name of the category can imply its constituents (just as in [Stansbury et al., 2013]'s example of the 'beach' category, which "is sufficient to elicit the recall of such objects as towels, umbrellas, sandcastles and so on.").

2.2 Data streams for complex event analytics

As I mention above, data on human understanding of hazards have been a focus of interest of behavioural geographers since the last century. However, it has been accepted that such data is hard to come by as compared with information about the physical extent and magnitude of natural hazards, as attitudes towards nature are strongly influenced by the socio-cultural backgrounds of event participants [Renn, 2008]. Nevertheless, the growing interest in approaching natural hazards as *socio-natural systems* has provoked interest in understanding how they are structured from the perspective of human behaviour as events people partake in and react to.

In order to reveal potentially successful routes of integration of behavioural and social data with other data streams, I looked at the broader areas of *complex event analytics*, dealing with the problems of *data ontologies* and *interoperability*. Since such smart cyber-physical systems (CPS) are based on processing of heterogeneous and dense data streams from multiple sensors, the main challenge turns out to be the integration and matching of patterns against those streams.

Since its inception, complex event analytics-related approaches have been adopted by operational intelligence and service management programmes in order to provide insights into business operations and/or identify actions that trigger particular events or their various states. They have been heavily based on *correlation*, which helped to analyze the majority of events and to select the most significant ones, so-called 'triggers'. This proto-complex event analytics did not produce new inferred events; instead it was used to compile event hierarchies (e.g., for relating high-level events with low-level events) and has subsequently evolved into business process management operations (BPMO), where seemingly disparate events significantly affect each other's outcomes (e.g., a malfunction or breakdown detection). Later, complex event analytics also started contributing to BPMO at implementation levels, such as customer service monitoring (e.g., click-stream analysis, customer experience management and 'recommendation systems'). Another early adopter of complex event analytics technology was the financial services industry, which used it for structuring and contextualising data streams to inform trading behaviour (e.g., algorithmic trading). More recent advances in complex event analytics technologies have also made it more affordable for smaller companies, thus enabling them to create their own algorithms to compete with larger players in spheres such as of online banking and multichannel marketing.

Whilst business and finance complex event analytics have been predominantly underpinned by *time series processing*, where historical time series and real-time streaming data are aggregated into a single continuum, more recent and systemic examples of complex event analytics can be seen in driverless cars and IoT, where sensors enabled detection of various events and trigger reactions that subsequently turned into events themselves (i.e., 'cascading events'). The majority of techniques are defined by their main property of *temporality*, i.e., a system's state is more efficiently presented in the form of dynamic data streams from which a system continuously learns a number of different states, rather than preserving its form as a static, materialized model.

More recently, [Wang and Cao, 2012] introduced *context-aware semantic* complex event analytics with the 'context' still being understood as metadata or annotation, however, this time containing more information on the background physical environment. Domain ontology was therefore used for the first time to annotate heterogeneous data and infer high-level contexts, accounting for physical entities, such as *Object* (things that participated in the event), *Time* (time instance or interval of the event), *Environment* (location and location features) and *Action* (the execution degree, method, tools of action). By using semantic and context-awareness technologies, complex event analytics was seen as being able to infer hidden information and improve event processing precision [Wang et al., 2016] by dynamically annotating heterogeneous data in streams to infer high-level context and subsequently associating detected events with the current context to forecast what will happen based on relationships in the event ontology.

Recent studies have demonstrated an increased interest in combining complex event analytics with social media monitoring (SMM), which would enable the possibility of early identification of potential users/customers or quantification of the influence of real-world events on public opinions, with the possibility to design systems capable of addressing cases with a strong aspect of *proactive* crowd behaviour (i.e., able to anticipate future behaviour in social media streams and act proactively to realize new opportunities or to avoid potential problems). Increased interest in *human-centric sensing* as a relatively cheap data source for various applications also provoked interest in testing these growing volumes of new data alongside quickly evolving complex event analytics approaches. As social media and other web data sources offered much richer semantic information volumes, emerging research started looking beyond mere metadata/annotation elements, such as, for example, time and geolocation, thus shifting methodological requirements towards *multimodal* analytical approaches [Thang-Duong et al., 2017].

2.3 Research framework

2.3.1 Reasoning behind data selection

The choice of social media data to study human perceptions of natural hazards was a straightforward one. People usually make an effort to stay aware of what happens in their neighbourhoods, especially if their health, wellbeing or prosperity is at stake, therefore a good part of their everyday experiences are likely to be ‘mediated’ by new communication tools. It has already been shown that public information sufficiency, risk perception and self-efficacy can predict risk-information seeking behaviour [Huurne and Gutteling, 2008; Krohne, 1989; Zuuren and Wolfs, 1991]. As a consequence, it goes without saying that natural disasters are extremely ‘newsworthy’ and, as well as actively ‘consuming’ risk-signaling reports, people also tend to engage in conversations about approaching, ongoing or past natural disasters on social media and various websites. There is a high degree of certainty that in any given year events such as storm, hurricane, earthquake, or even volcanic eruption will steal the headlines; and when they do, it is equally certain that the focus of attention will be stories of immediate public interest, including the havoc and disruption of everyday life, destruction of houses, numbers of victims, and first person reports of witnesses.

Developments associated with Web 2.0 have exponentially increased the amount of digital information, predominantly textual, which, when collated as data elements, can be used as standalone information signals or accompanying descriptors for acoustic, photographic and video material. Also known as a source of ‘natural language’ data, these sources of unstructured information are very well positioned to help understanding how meanings are created and to expand applications towards event monitoring.

Grounded in geocomputation, citizen science is increasingly seen as a nexus at the interface of policy, science and the public [Newman et al., 2012], and various types of similarly produced user generated content (UGC) have been employed in ecological and environmental studies in order to explore the opportunities they afford to quantify relationships between physical phenomena and community response, as well as to assess the reliability of citizen science data in relation to expert data [Riesch and Potter, 2013]. Design- and principle-wise, these novel and emerging data sources vary from *secondary* (customizable multi-purpose *templates*) and *primary* (Facebook, Geocaching, etc.) web platforms to custom-made smartphone apps (e.g., RiverObstacles) able to accommodate various types of data collection, ranging from text and numerical and from passive to active sensing. In most cases, such practices have proven to be an economic and efficient alternative to data collection for scientific purposes, where results indicated high concordance between participants’ and expert scores [Riesch and Potter, 2013].

Known ongoing challenges for citizen science initiatives include wider demographic engagement with such initiatives, as well as the need to ensure continuity of data collection routines. Thus, whilst it is relatively easy to find individuals willing to share information using tools provided by Web 2.0, it is not always obvious how to ensure ongoing engagement from volunteers, especially for cases where active and continuous sensing is required [Silvertown, 2009]. Also, from the data standards perspective, as a collaborative outcome and because crowdsourced information is often analyzed in terms of the ‘big’ data sets, the lack of the possibility of granular tracking of all data points brings forward concerns about data quality and mistrust of citizen science in some circles of the scientific community [Lopez-Aparicio et al., 2017]. This issue is specifically noticeable for cases where data is often unstructured and/or generated opportunistically. The issue is less apparent for cases with a structured format (like data collected via dedicated platforms and smartphone apps), which makes curation somewhat easier. [Riesch and Potter, 2013] also point out that the most challenging aspect in environmental citizen science user generated content (UGC) data is the lack of complete and accurate geolocation data associated with descriptive information. This subset of UGC, also known as Volunteered Geographic Information (VGI), is considered the most promising domain of citizen science, however, it also suffers from data collection disparities due to unequal access to Information Communication Technology (ICT) by different socio-economic and demographic groups, creating forward ontological problems in data production, which, however, do not necessarily affect its status or quality [Silvertown, 2009].

New data and new ways of working with these datasets therefore suggest more creative planning methods capable of equally incorporating readings from distributed sensor networks and interpreting various social participatory undertakings [Townsend, 2013]. Some of them are strictly design-related. For example, the design of intelligent monitoring of social media with proactive event detection properties [Riemer et al., 2012] also introduced a set of challenges related to real-time data processing, such as online semantic indexing used in contemporary web searches. Semantic capabilities have been defined as one of the crucial components for event pattern definition (as opposed to simple filtering) that allow users to express their interest on a more abstract level, thus simplifying *user-centric* pattern generation and allowing usage of semantic requests based on previously specified domain knowledge in event processing engines.

Since the majority of socio-environmental processes are *spatially grounded*, geo-referenced social media data is considered to be fundamental for complex event analytics routines, specifically for projects with environmental applications. Current methods are primarily orientated towards mining vast linguistic resources in their natural state without taking into account its own dynamics, which could be expressed, for example, in form of the lexico-grammatical, phonological and sociolinguistic determinants of language production. As a consequence, vast amounts of potentially useful semantic information on the social web continues to be accumulated in its latent form.

Since human communication can acquire both verbal and nonverbal forms, a criticism has been already been expressed in recent research overviews [Yang and Eisenstein, 2013; Eisenstein, 2017] that the field of orthodox computational linguistics is overly concerned with "the structure of verbal information transfer", whilst ever increasing availability of social media data opens opportunities to far more challenging interdisciplinary methodological undertakings, such as better understanding of language settings and all kinds of social processes. [Nguyen, 2017], for example go as far as describing social media data as "a data type that is signaling all kinds of social phenomena", and provide some convincing arguments for the obvious need to formalize this methodological sub-domain of computational linguistics into a separate subfield of the computational sociolinguistics that builds its methodological capacity on the connection between social variables and language used by the same socio-demographic groups. This connection is primarily grounded in linguistic variation due to colloquial nature of human interaction on social media; and although it can be characterized as fluid and tenuous, it, nevertheless, captures the symbolic nature of the socio-linguistic dualism by means of representing the speaker's social identity via the language devices they choose to use in the particular circumstances.

An example of this successful methodological formalization provoked me to reflect on where social media stands it terms of complex event analytics requirements in general and, particularly, how it can be re-purposed to answer the research questions of this thesis. Some further research into this aspect has demonstrated the peculiarity of the situation as, on the one hand, social media data and event processing intuitively seem to be two complementary elements of the same system due to the inherently event-driven nature of the data production on digital social platforms. This requirement has therefore motivated me to select a platform for experimental analyses that includes several compulsory modalities (e.g., annotation tags/text, visual material that can be represented by photographic or videographic materials, geolocation/coordinates, etc.). For this purpose, I chose Yahoo Flickr Creative Commons 100 Million dataset (YFCC100M) [Thomee et al., 2016], which is a global-coverage open dataset provided by Yahoo! for the research community to experiment with fully multimodal crowd-generated source of information and comprising the main modalities of my interest: *linguistic* (tags), *geographical* and *visual* (Fig. 2.1).

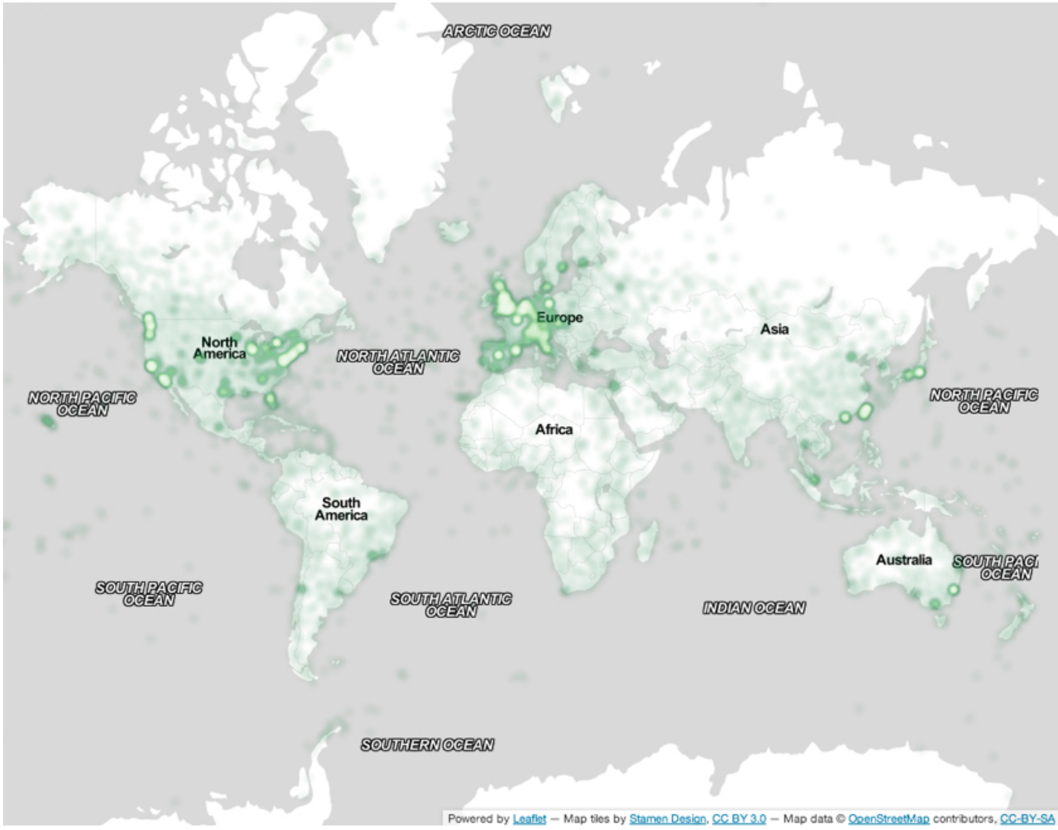

Figure 2.1: Flickr Creative Commons dataset is the part of Yahoo Webscope's datasets for researchers, and is considered to be one of the largest public multimedia datasets that has ever been released (99.3M images (49M of which are geotagged) and 0.7M videos), all from the single platform and released under Creative Commons licensing. The dataset (appr. 12GB) consists of such attributes as *photo ID*, a *jpeg url* or *video url*, and some corresponding metadata such as *title, description, camera type* and *tags*. Other attributes, which are derivatives of the Flickr social ecosystem, like *comments, favourites* and *social network data* can be queried from the Flickr APIs directly. *Image by David Shamma on Flickr (CC BY-ND 2.0).*

2.3.2 Data collation

When trying to associate data with a particular event, we use domain-specific keywords in order to describe those events (i.e., 'weather', 'rain' and 'storm' for natural hazards analytics), which are subsequently related to the locations they are, or likely to, originate from. Therefore location and keyword descriptors are of the upmost importance for natural hazard analytics from the phenomenological perspective – however, both have some known limitations due to their *volunteered* nature as data signals.

When we take Twitter as an example, geographic metadata available for geo-located tweets predominantly originates from profile location (82 per cent), mentioned location (17 per cent) and activity location (1 per cent) [GNI, 2013]. Since the particular value of data streams for flood monitoring purposes lies in their capacity to generate large volumes of georeferenced data, where social media data is concerned as a potential contributory source, its entry points should consist of the point-lexemes with the precise XY coordinates as opposed to the place names, which are less accurate due to their nature of being topological centroids of the real world objects. The fact that our data derives from social media platforms suggests that the data coverage will be uneven, and as a consequence this will find reflection in the spatial component of our multi-modal dataset, since:

(i) people, living or visiting flood risk areas are not necessarily representative of the demographics of social media or mobile technology users,

(ii) and even if they are, they do not necessarily have geolocation enabled on their smartphone/mobile devices,

(iii) or are not actively broadcasting the topic of flooding on their profiles.

Hence there is a need to develop alternative methods for precise location extraction from social media data.

Keywords, which are widely used as direct event descriptors, also have their limitations as prospective data sources. Although analytic philosophy still claims that any meaning can be expressed in *language* [Searle, 1979], outside of it the limits of natural language when it comes to meaning-making have long been recognized in both arts and sciences. Psychology and linguistics acknowledge that language is not a perfect medium as much of our thought is either ambiguous or inexpressible in language, so there is scope to look for new data sources - or their combinations - and algorithms, capable of highlighting any deficiencies in natural language and adapting it to specific pragmatic purposes, such as, for instance, environmental sensing and risk-signaling.

2.3.3 Social media data for natural hazard analytics

Since human experience of environmental processes can be ambiguous, we are interested in the detailed insights granular data from social media can provide about meaning creation and the dynamics people attribute to their surroundings whilst experiencing a natural hazard, or any risky situation for that matter. Hence, the two main mechanisms of improving natural hazard monitoring with help of social media data considered in the scope of this thesis are: **(i)** *useful volume* and **(ii)** *semantic plasticity*, where it can be argued that in the context of our primary data source, both elements are interrelated and interdependent.

As the mechanism of useful volume has been already briefly covered here and elsewhere as a known data instrument, below I am going to provide a brief explanation of the semantic plasticity mechanism, which has not yet received any attention in the research literature as a potential tool for event analytics.

Semantic variation of form is considered as a fundamental language property of *plasticity*, inherent to its structure and driven by universal cognitive mechanisms that are explained by a dynamic conception of meaning construal, and hence possess mechanisms of self-regulation expressed in complementary rules of *variation* and *stability* [Robert, 2008]. For instance, various metaphorical and metonymic forms are known to result from a certain kind of stability governed by familiar mechanisms of common meanings and prototypes. Whilst this plasticity is often discussed in linguistic circles from the perspective of *ambiguity* and as a source of communicative misunderstandings, it nevertheless can be also examined from the point of view of dynamic lexical data resources, which can be exploited by various event tracking methods, as linguistic references are always mediated (or grounded in objective reality). It is very rare that form and its meaning have a unique one-to-one correspondence, usually a form possesses several meanings, which can be construed in accordance with the context and using various mechanisms, such as categorization or segmentation of phenomena, selection or highlighting of event specific properties, etc., since reality is presented to perception as a continuum to be labelled by discrete linguistic resources. Construction of varying designators of the referent also explains the existence of polysemy and synonymy, which can vary from language to language.

Back in the late 80s to the early 2000s, when mechanisms behind the brain's handling of meanings were not well known [Pulvermüller, 2001], cognitive linguists speculated about the existence of so-called image schemas, which constitute a form of representation that is common to perception, memory and semantic meaning [Gärdenfors, 2007]. In their work on spatial prepositional structures, [Lakoff, 1987; Langacker, 1987] proposed the system of the schemata constituents, which, for example, can comprise the *trajector* (the object that is in focus) and the *landmark* (in relation to which the trajector executes its spatial functions of movement or positioning), and therefore is universal and non-descriptive. [Gärdenfors, 2007] also points out the strong connection between image schemas and visual processes, where 'trajectory-landmark' duality corresponds to 'figure-background' in the field of visual perception and both trajectory and figure are in the focus of attention. According to [Lakoff, 1987; Langacker, 1987; Johnson, 1987], the two main axioms of image schemas are: **(i)** their structures are inherently spatial, and **(ii)** they are universal, i.e., every word is supposed to be part of an image schema. Also, according to [Langacker, 1987], a schema can be re-purposed into another one with the same dimensions, objects and relations by changing only its *focus*. This is when the process of *refocusing* takes place, where the same scene can serve as a platform for very different cognitive processes involved into processing of its different aspects.

In his attempts to formalize image schemas for computer processing applications, [Holmqvist, 1993, p.31] defined them as "that part of a picture which remains when all the structure is removed from the picture, except for that which belongs to a single morpheme, a sentence or a piece of text in a linguistic description of a picture." [Gibbs and Colston, 1995, p.348] define them as "dynamic analog representations of spatial relations and movements in space."

In his earlier work [Gärdenfors, 2000] presented a precise theoretical account of what constitutes an image schema based on the notion of *conceptual spaces*, which had been used by two distinct cognitivist traditions in linguistics. The first one, represented by Lakoff, Langacker and Johnson defined spaces as the platforms or frameworks for domain modeling, onto which can be mapped various spatial and temporal dimensions, and hence can be regarded as mere geometrical or topological structures. [Lakoff, 1987] expressed the claim that meanings of linguistic expressions should be understood through the lens of spatial image schemas combined with the metaphor overlay.

Unlike Lakoff and Langacker, the second approach brings into focus the *dynamics of representations*, rather than geometry of image schemas, which therefore lose their definition of 'images' as such. The early proponent for this approach is [Talmy, 1988] who described the importance of the role of underlying forces and dynamic patterns, which could be implied from visibly static scenes and hence are better positioned for meaning extraction. Followed up by [Barsalou, 1999], the concepts were redefined as "perceptual symbols that are dynamic patterns of neurons functioning as simulators that combine with other processes to create conceptual meaning."(p.611) In Barsalou's theory, these meaning containers are very closely related to perceptual processes.

Since this thesis is concerned with the problem of *activation* of semantic resources in unstructured ('natural') language data for an applied task of flood risk monitoring, it seems inevitable to look into problems of meaning dynamics, specifically what drives its change, duration of change and under which circumstances.

2.3.4 Research avenues

As it has been outlined in the literature review, there are currently several outstanding research avenues that could be followed up with help of the various modalities of social media data available for collection via numerous API instruments. Some prominent examples have already been investigated during the data exploratory stage of this project and are presented in Table 2.1, and specifically relate to questions of public community engagement strategies, natural hazard risk perceptions and eyewitness reporting of disaster impacts. Some of those and very similar research problems have already been addressed with help of more traditional data sources and methods, predominantly of qualitative and interpretative traditions of social sciences. Their subsequent validation with social media analytics was mainly motivated by opportunities these emerging data sources had to offer in term of cost, volume and flexibility as the same databases could (and still can) be re-purposed to address very similar or slightly different sets of research questions.

The use of social media in natural hazard analytics specifically has been already widely covered in the recent research literature [Earle, 2010; Acar and Muraki, 2011; Preis et al., 2013; Al-Saggaf and Simmons, 2015; Tang et al., 2015; Kaufhold and Reuter, 2016; Crooks et al., 2013]. Unsurprisingly, the majority of studies were concerned with use of VGI (Volunteered Geographical Information) component of UGC due to the strong interest from the research community in the 'sensing' properties of social media signals, which could be useful for hazard monitoring in poorly or un-instrumented locations. The authors of this research tradition of mixed methods strongly advocate the combined use of signal streams from social media platforms and authoritative physical sensors, where the former are regarded as suppliers of more intense (yet noisy) data complementary to official instruments that are usually with limited geographical coverage, but more trustworthy scientifically.

If we take look at the current gaps in the domain of natural hazard analytics with help of social media data, it become obvious very quickly that there is a certain lack of 'applied critical' studies that could, for example, either evaluate the full potential of one single platform for a range of questions in particular domain of natural hazard analytics or compare the results across several social media platforms. Another observable criticism is that, despite almost a decade of social media analytics, there still exists a certain degree of mistrust and skepticism in scientific circles regarding use of social media data on its own, hence the proliferation of analyses where social posts are converted into 'signals' and merged with the readings of physical sensors. Also, if we recall again the problem of the lack of a phenomenological perspective in social media studies predominantly concerned with dynamic social processes, it is therefore possible to conclude that there exists a certain degree of disciplinary discrepancy regarding how the same data source is being approached by the different (here: social and environmental) disciplines. Since the primary purpose of this thesis was to advocate these methods for advancing understanding of natural disasters from perspective of human linguistic and behavioural interpretations, it turns into an interesting challenge to demonstrate how this discrepancy can be approached empirically.

	Profile	**Text**	**Tags**	**Geolocation**	**Timestamps**	**Visual media**	**Followers**	**'Social buttons'**	**Comments**
1. Natural hazard protection (ecosystem service)		[Tkachenko et al., 2015]	[Tkachenko et al., 2015]					[Tkachenko et al., 2015]	[Tkachenko et al., 2015]
2. Public engagement	[Tkachenko et al., 2016a]	[Tkachenko et al., 2016a]	[Tkachenko et al., 2016a]	[Tkachenko et al., 2016a]			[Tkachenko et al., 2016a]		
3. Pre-event estimation of the natural hazard effects				[Tkachenko et al., 2016b]	[Tkachenko et al., 2016b]				
4. Estimating natural hazard effects on social media and satellite images			[Tkachenko et al., 2017c]	[Tkachenko et al., 2017c]	[Tkachenko et al., 2017c]	[Tkachenko et al., 2017c]			
5. Post-event estimation of the natural hazard effects			[Tkachenko et al., 2017b]	[Tkachenko et al., 2017b]	[Tkachenko et al., 2017b]				

Table 2.1: Preliminary studies of the human perception of and response to the natural hazards with use of various components of the platform-enabled (i.e., API-configured) social media data.

Chapter 3

Methodology

3.1 Synthesis

In the scope of this Chapter I present the main arguments behind the choice of the research questions selected for further empirical verification. I start off by presenting semantic drift as a main analytical tool for analysis of the human perception of flooding on the multimodal platform Yahoo! Flickr. This discourse is followed by hypotheses theoretically based on state-of-the-art approaches in the field of distributional semantics and is repurposed for further practical verification in context of the flood events. The second half of the Chapter is dedicated to the theoretical reasoning behind selection of the *primary* (e.g., variables-candidates for semantic drift) and *secondary* (authoritative flood monitoring datasets) data sources for the three research questions of this thesis.

3.2 Background

3.2.1 Research questions

Following the arguments outlined in the literature review, the main reasons behind the choice to research natural hazard events (specifically floods) with the help of semantic change on social media are:

(i) To find out whether event-specific (as opposed to known sociocultural long-term irreversible semantic changes of interest to the field of diachronic linguistics) language dynamics exist on social media;

(ii) To verify whether semantic drift can increase the volume of useful event-specific georeferenced data as compared to the dominated use of the direct event descriptors. The notion of *usefulness* is defined here in the context of outstanding problems in flood risk management, notably, problems of **detection** (*can alternative data sources 'sense' event outbreaks?*), **differentiation** (*do communities understand what type of flooding they are exposed to?*) and **segmentation** (*how does annotated visual material represent people's perceptions at different stages of flood events?*).

My secondary research interests are linked to questions of critical platform evaluation, where I want to use a single platform across several interlinked research tasks.

3.2.2 Semantic drift as an analytical tool

There are two main reasons why semantic drift was selected as an analytical tool, one theoretical and one pragmatic.

From the theoretical perspective, there is an emerging interest in verifying how events people experience or take part in are reflected in the data footprints they leave on social media platforms. Following from here, there is also a more practical interest in verifying whether changing meaning dynamics affects the amount of useful data that can be re-purposed for better event monitoring, management and intervention.

The first studies of semantic change emerged at the end of the 18th century, with the most famous one conducted by [Bréal, 1897], who coined the term 'polysemy', a property of linguistic units (e.g., words) to acquire several meanings and therefore to be more prone to semantic change than words with fewer meanings. Usually in the linguistic literature *polysemy* is described alongside synonymy, where they tend to be opposed to each other as the functional-conceptual variation of any symbolic form to the functional-conceptual relation between any symbolic forms. Seen in these terms, both synonymy and polysemy can be described as either concrete or schematic. Thus, [Glynn and Robinson, 2014] cite Lakoff's analyses [Lakoff, 1987] of spatial prepositions as studies of *concrete*, or non-schematic, polysemy, whilst his study of the deictic construction is regarded as *schematic* polysemy.

Following the earliest examples of studies of polysemy-synonymy, which were predominantly quantitative, [Geeraerts, 1993; Wierzbicka, 1990; Taylor, 1995; Hilferty, 2015] also highlighted theoretical links between the two concepts, thus leading to discussions on semantics of categorization in cognitive linguistics. In these discussions, semantic ambiguity of polysemous words was defined either by the number of interpretations [Rodd et al., 2002] or by their relationships that define their ambiguous conditions [Klepousniotou and Baum, 2007]. While the first approach was predominantly concerned with dictionary studies, the second one paved the way towards redefining 'polysemy-synonymy' relationships from the perspective of *categorical hierarchies*, where a polysemous term became known as 'a prototype' and the associated synonyms as its synset formations. Subsequently, [Klepousniotou and Baum, 2007], in further investigations into relational polysemy tested recognition of figures of speech, metaphors and metonyms in the context of their *regularity* and came up with the conclusion that relationships between the metonym senses are more likely to be regular than relationships between senses of metaphorical words, and they also are predicted to be processed faster than metaphors. The latter findings have been challenged by [Jager and Cleland, 2015] as they suspected that in the original [Klepousniotou and Baum, 2007] study the number of senses were not controlled for and their effects were confounded by numerical polysemy.

The starting point of this approach was the realisation that it is impossible to describe every aspect of the real world in terms of a delimited number of semantic components and we should look for some kind of 'ideal' category member (a.k.a. prototype) on the basis of semantic proximity to other members, which, alongside the prototype, are composed into a *semantic category.* According to some followers of prototype semantics [Labov, 1984], category boundaries may shift depending on the linguistic or situational context. According to other authors [Klepousniotou et al., 2008; Kacinik and Chiarello, 2007; Jager and Cleland, 2015] the more members (a.k.a. synonyms, or *synsets*) are grouped around the prototype into the category, the more *polysemous* (or ambiguous) the prototype is.

Although it has been already recognised that there was less work on synonymy per se, conceptual metaphor and metonymy studies have been, in effect, synonymy studies. Such studies were primarily focused on figurative lexemes and, according to some authors [Kittay and Lehrer, 1981], they look like studies of 'near-synonymy'. Obviously, they included much discussion on what constitutes *a source* or *a target* domains and whether certain expressions represented the concept in question well. All these and similar questions are concerned from a lexical semantic point of view with the problems of [near]synonymy.

The main descriptive models of polysemy used by the cognitive linguists are therefore *prototypicality* and *semantic frequency*. The role of lexemic frequency has been originally discussed by [Hamilton et al., 2016b] in the context of the *Conformity Law of semantic change*, where it was argued that words that are being used most frequently are less prone to semantic change as their main function is to support everyday communication. Similarly, according to the *Law of Prototypicality* [Dubossarsky et al., 2017], words and lexemes that are used as descriptors of a particular lexical category with a degree of centrality in the radial system, i.e., are either central or near-central members, are also less likely to drift semantically due to their core function of categorical support. Finally, according to the *Law of Innovation*, polysemy itself is positively correlated with semantic change [Hamilton et al., 2016b].

[Hamilton et al., 2016b]'s findings are challenged in [Dubossarsky et al., 2017], and examining the *Dubossarsky-Hamilton dilemma* in more detail, it turns out that factors leading to semantic change are far more diverse and not necessarily limited to purely distributional (i.e., linguistic) ones. For example, an initially proposed negative correlation between semantic change and word frequency has been shown in [Dubossarsky et al., 2017] to occur mainly due to choice of models and its role, although contributory, is not as significant as originally suggested by [Hamilton et al., 2016b]. The proposed negative correlation between meaning change and *prototypicality*, as well as a positive correlation between meaning change and *polysemy* turned out to be much weaker as both are highly collinear with frequency and therefore cannot be regarded as independent contributors to semantic change. It was proposed therefore to look at some other, additional, factors from which meaning change may result and it is most likely due to their interaction that we can observe, according to [Dubossarsky et al., 2017], likely more credible effects, however small they turn out to be.

Such factors may originate from various *situational* domains, notably social, political or environmental [Bochkarev et al., 2015; Newman, 2010] and they require some further investigation. For example, [Kulkarni et al., 2015] argue that linguistic shifts are especially typical for the language of the Internet, which is more likely to stimulate or provoke meanings changes in the context of the rapid exchange of ideas. They used several methods, notably frequency-based and the distributional ones in order to construct word time series and then model temporal evolution of natural language and observed that initial choice of the construction method can determine types of extractable information about word usage trends. Specifically, they observed differences between the words Sandy and Hurricane on the Google Trends platform in October 2012, where both demonstrated well-defined frequency spikes and therefore were expected to undergo the shift in meaning (according to the authors' assumption). However, only the word Sandy had actually acquired new meaning, which was revealed with help of the distributional semantics method. It has been therefore concluded that the methodological value of the *frequency* analysis can be considered as an initial step in assisting selection and identification of potential lexical candidates for semantic change.

According to [Bochkarev et al., 2015], the dynamics of lexical evolution can be essentially of two types. The first one is the general lexicon contingency of historical factors, which is fairly uniform across multiple languages and at timescales of at least five decades (i.e., *macrochange*) and the second one is driven by regularities, possibly universals of cognition and social interactions (i.e., *microchange*), which are variable and differ between languages as they can be driven by societal transformations or catastrophic events.

3.2.3 Hypotheses

Following conclusions of the abovementioned authors, in order to maximise extraction of lexical units covering a particular meaning of our interest (e.g., *flooding*), we can use the following reflections.

First of all, it would be fair to assume that *useful* data expansion will take place at the expense of **higher-frequency synonyms** of the direct word-descriptor of the event in mind and/or its **prototype**. The direct event descriptor itself can be either one of the synonyms in the cluster of synsets (sets of synonyms), belonging to the particular category or a prototype itself. If the latter is the case, we need to ensure that the prototype, being essentially polysemous by nature, is related exclusively to the event of our interest and does not signify any other phenomenon. One way of dealing with this issue experimentally is to ensure that it is strongly related to the least polysemous synset(s), which also represent the potential candidates for *useful* semantic change, and this strength can be confirmed either ontologically or topically.

For the similar reasons, situations where the direct descriptor is a prototype are also highly undesirable, because it will have a higher probability of drifting semantically itself, hence reducing the data volume in the opposite direction, so we need to consider cases as optimal where the direct event descriptor belongs to one of the synsets. Data expansion in such cases is therefore possible if the candidates for semantic change relate to the direct event descriptor in the capacity of a *parallel synset*. However, if the direct event descriptor is either a prototype of the category or one of its [highly] polysemous synsets itself, we need to consider the introduction of some kind of a 'lexical benchmark', i.e., a word or set of words that are strongly connected (either ontologically or topically) with the candidates for semantic drift - but not with the direct event descriptors. Hypothetically, multidirectional drifts are simultaneously incompatible, so the drift towards a different category would mean that the meaning candidate data is void in terms of its usefulness for our complex event analytics. However, this hypothesis requires further independent empirical testing and is beyond the scope of this thesis.

When we shift our attention to the physical forms in which meanings tend to manifest themselves in any language, we are inevitably going to hit the domain of *usage-based linguistics*, which claims that meanings are operationalised either via *pure monosemic* or via various manifestations of *polysemic* forms (retrospective equations, where a single item has multiple meanings with equal etymologies). Multiple meanings (synonyms, or synsets as it has been introduced above) of the latter can relate to each other as metaphors (i.e., *extrospective equations*) or metonymies (i.e., *introspective equations*).

It can be argued that metaphors are potentially *less useful* for the purpose of this thesis as due to their frequent irregularities they may not cover sufficient amount of data to render them *useful*. Nevertheless, metaphors still represent an interesting candidate data, especially in the context of prototype theory as they can represent particular instances of measurement, quantity and comparison, and therefore be useful for *typological* complex event analytics. When things or objects do not have clearly defined physical or conceptual boundaries, such as natural hazards, we still have a tendency to classify or categorise them as such, because we are inclined to perform mental operations with discrete objects, taking as an example "mountains, street corners or hedges" [Lakoff and Johnson, 1980, p.25]. Such way of perceiving the physical world is usually motivated by certain practical purposes we have on mind, such as "locating mountains, meeting at street corners, trimming hedges", or even vital ones, such as risk detection and self-preservation. Similar to his work with prepositional radial structures, which illustrate basic experiences of human spatial orientation and producing *orientational* metaphors, [Lakoff and Johnson, 1980], operationalisation of the concepts behind physical objects provides the basis for *ontological* metaphors, which allows for various and alternative ways of seeing events and phenomena in much more compartmentalized ways that are easier to perceive and understand. Interestingly and understandably, as [Lakoff and Johnson, 1980] point out, very rare cases of ontological metaphors are seen as such in classical encyclopedic understanding of the term as they serve a very limited range of purposes – such as referring or quantifying. For example, both *spate* and *deluge* can transfer their metaphorical sense of measure onto flood phenomena, where the first one originally signified a large number of similar things coming in quick succession and the second designated a great quantity of something arriving at the same time, but in the context of natural hazards they, for example, can mean different types of flooding, a surface water/riverine and the pluvial/flash floods, respectively.

3.2.4 Prospective applications

In the above sections I predominantly concentrated on linguistic rules of semantic change and how they can be linked to the topic of natural hazards theoretically. However, in order to render them *useful* for the purpose of event analytics, in addition to the increase of useful data signals, it also should do so in agreement with the specific problem tasks of flood risk monitoring and management, which have been defined in the scope of this thesis as *detection*, *differentiation* and *segmentation*. These specific problem tasks are related, for example, to the outstanding challenges in design of more *socially inclusive* flood risk communication programs, to understanding of public reaction to the current authoritative warnings and to appreciation of the ways social media is able reflect crowd responses to the different types of flood events.

The decision to work with a single social media platform was motivated by interest in verifying what is the maximum potential it holds in covering several aspects of event analytics, however, one important condition here was its strict multimodality, i.e., each text message should be associated with other types of crowd-sourced data elements or their metadata, such as timestamps, geolocation and visual media, where in our case the latter is the primary data component of the photo-and videographic Yahoo! Flickr platform. Table 3.1 presents the structure of the primary and supporting data components, derived from the Yahoo! Flickr platform in order to construct individual models to answer the research questions. The exact configuration of the models will be expanded in their respective following chapters.

Research question	Primary data elements	Supporting data elements	Proposed model
1. *Can we predict events (e.g., flooding) with help of semantic drift on social media?*	Tags, timestamps	Geolocation	Transient semantic microchange (Chapter 4)
2. *Can semantic drift on social media help to differentiate types of flood events?*	Geolocation, tags		Spatial semantic change (Chapter 5)
3. *Can alternative tags help to distinguish different stages of flood events?*	Tags, images	Timestamps, geolocation	Cognitive semantic drift (Chapter 6)

Table 3.1: Proposed models of event-driven semantic change, re-purposed for the analysis of mediated human behaviour during flood events.

3.3 Data components

3.3.1 Selection of the lexical variables

According the literature summarised above, the following conditions need to be satisfied by candidate variables for semantic drift in order to be regarded as *useful* data for event analytics:

(i) They should have a higher frequency than direct event descriptors in order to be able to provide additional data signals. According to the Conformity Law of semantic change [Hamilton et al., 2016b], word frequency correlates negatively with change of meanings, however, since this condition was challenged by some other authors [Dubossarsky et al., 2017] and frequency itself is an important factor for definition of *useful* data, here it was regarded as a factor that requires further verification in conditions of situational semantic change (or *microchange*).

(ii) They should exhibit a more *negative* sentiment score as compared with a direct event descriptor in order to have increased semantic instability and hence are more likely to drift semantically. Following Hamilton's work on semantic drift [Hamilton et al., 2016b,a], the implications are that negative words are more diachronically unstable and therefore they have faster rates of semantic change.

(iii) They should be *polysemous* (i.e., have a relatively high number of synsets). It has been agreed between several authors that polysemy correlates with semantic drift, however, where event analytics is concerned with data maximization, we want to avoid situations of 'multidrifts', where both direct event descriptors and candidates for *useful* semantic drift exhibit fluctuations. This condition also requires further verification in the context of event-focused analytics.

(iv) They should have high degree of *prototypicality* (several strong relationships of 'is-a' type with their synsets), this condition is directly related to polysemy, hence it also requires further adaptation for studies of event-driven semantic instabilities.

All these conditions have been a focus of attention in recent literature [Dubossarsky et al., 2017; Hamilton et al., 2016b] concerned with the rules (or laws) of long-term sociolinguistic types of semantic changes. Since no such laws, to the best of my knowledge, have been proposed for the short term, situational semantic drifts, I therefore re-purposed the above-mentioned conditions for my case of event analytics in order to verify whether they may also applicable for the cases of phenomenological semantic instabilities and rendered useful from the practical point of view.

Since event analytics presumes *context similarity* between direct event descriptors and candidates for semantic change I therefore I started my data selection by matching my principal single lexeme 'flood'(F) with the words that have high *semantic similarity* (prototype test) and *relatedness* (test for synonymy) scores and are represented by higher volumes in the Yahoo! Flickr database (Fig 3.1).

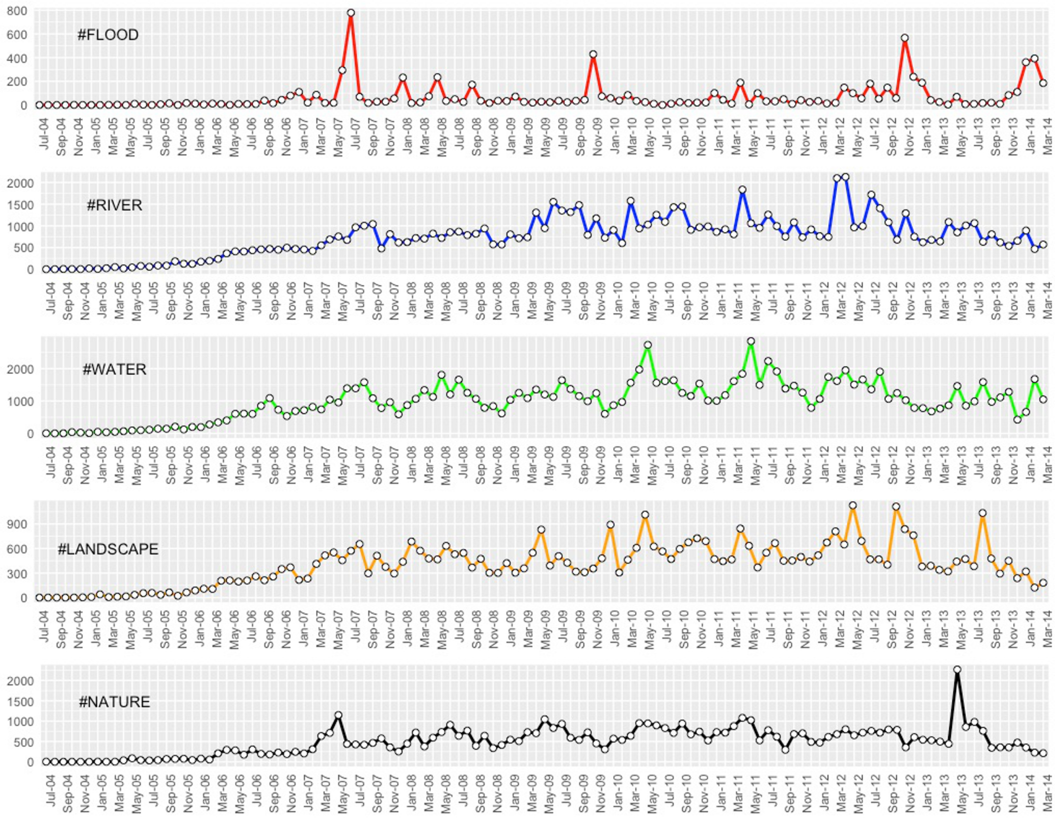

Figure 3.1: Monthly-aggregated temporal profiles of the main environmental tags (georeferenced) on Yahoo! Flickr (2004-2014).

Three *similarity* measures ('is-a' hierarchical type relations) used are based on the lengths of paths between the concepts: LCH [Leacock and Chodorow, 1998], WUP [Wu and Palmer, 1994] and PATH proper:

The three remaining *similarity* measures (RES [Resnik, 1970], LIN [Lin, 1998], and JCN [Jiang and Conrath, 1997]) are based on information content, which is a corpus-based measure of the specificity of a concept:

There are also two *relatedness* measures (e.g., 'has-part', 'is-made-of', 'is-an-attribute-of', i.e., non-hierarchical type relations) HSO [Hirst and St-Onge, 1998] and LESK [Banerjee and Pedersen, 2003]. The LESK measure uses the text of a gloss as a unique representation for the underlying concept and assigns relatedness by finding and scoring overlaps between the glosses of the two concepts.

The words that satisfied both conditions are lexemes 'river'(R) and 'water'(W), which are represented in the WordNet 3.1 database [WN, 2015] as mono- (having a single gloss) and polysemous (having several glosses) words, respectively:

'river' (R) gloss

(n) river (a large natural stream of water (larger than a creek)) "the river was navigable for 50 miles"

'water' (W) glosses

(n) water, H2O (binary compound that occurs at room temperature as a clear colorless odorless tasteless liquid; freezes into ice below 0 degrees centigrade and boils above 100 degrees centigrade; widely used as a solvent)

(n) body of water, water (the part of the earth's surface covered with water (such as a river or lake or ocean)) "they invaded our territorial waters"; "they were sitting by the water's edge"

(n) water (once thought to be one of four elements composing the universe (Empedocles))

(n) water system, water supply, water (a facility that provides a source of water) "the town debated the purification of the water supply"; "first you have to cut off the water"

(n) urine, piss, pee, piddle, weewee, water (liquid excretory product) "there was blood in his urine"; "the child had to make water"

(n) water (a liquid necessary for the life of most animals and plants) "he asked for a drink of water"

The structure of WordNet 3.1 is based on the principles of a differential theory of lexical semantics [Miller et al., 2008] and where representations are not on the level of individual words or word forms, but on the level of word meanings (lexemes). These lexemes are characterized by listing their word forms in *a synset* so the meanings of words and word forms are strongly influenced by sets of words they share meanings with. The meaning of a concept is hence determined by its position relative to other words in the larger database structure.

Going back to our case, in topological terms, data expansion here is therefore hypothetically possible at the expense of the lexeme (R) constituting parallel synset relations (as metonyms: 'is-part-of' type) with the word (F) and at the expense of the lexeme (W) entering prototypical (hierarchical 'is-a' type) relationships with both synsets (F) and (R) of a certain hydrologically themed category (unspecified in the context of this thesis as we are not exploring the entire category). And according to our earlier conditions outlined in the Hypotheses section of this Chapter, since one of our candidates for semantic drift is the prototype (and hence is essentially polysemous by nature), we need to ensure that it is related *exclusively* to the event of interest and does not signify any other phenomenon. This condition is satisfied by the presence of the monosemous lexeme (R) with which (W) is strongly connected both ontologically and topically, so if (R) is drifting semantically during flood event, the (W) lexeme is very likely to exhibit similar behavior under the similar circumstances.

According to my prior hypotheses, situations where the direct descriptor is a prototype are also highly undesirable because it will have a higher probability to drift semantically itself, thus reducing the data volume in the opposite direction, so we need to consider cases as optimal where the direct event descriptor is one of the synsets. However, if the direct event descriptor is a [highly] polysemous synset, we still need to consider the introduction of some kind of a *lexical benchmark*, i.e., a word or set of words that are strongly connected (ontologically, topically or both) with the candidates for semantic drift, but not with the direct event descriptors. This benchmarking will introduce additional point(s) of reference for situations of simultaneous 'multidrifts', so these additional variables preferably are less polysemous (if monosemy is unattainable) and have positive sentiment/connotation score, both conditions for relative sematic stability. In our case, lexeme (F) is relatively polysemous (six glosses, see below) and have a negative connotation (-0.03125 sentiment score [ASA, 2016]), both individually being sufficient conditions for semantic drift [Iliev et al., 2016; Hamilton et al., 2016a]. Using the previously mentioned principle of topological similarity ('is-a' type relationships) I selected two benchmark lexemes 'nature' (N) (five glosses; 0.25 sentiment score) and 'landscape' (L) (four glosses; -0.015625 sentiment score).

'flood' (F) glosses

(n) flood, inundation, deluge, alluvion (the rising of a body of water and its overflowing onto normally dry land) "plains fertilized by annual inundations"

(n) flood, inundation, deluge, torrent (an overwhelming number or amount) "a flood of requests"; "a torrent of abuse"

(n) flood, floodlight, flood lamp, photoflood (light that is a source of artificial illumination having a broad beam; used in photography)

(n) flood, overflow, outpouring (a large flow)

(n) flood, flowage (the act of flooding; filling to overflowing)

(n) flood tide, flood, rising tide (the occurrence of incoming water (between a low tide and the following high tide)) "a tide in the affairs of men which, taken at the flood, leads on to fortune" (Shakespeare) 'nature' (N) glosses

(n) nature (the essential qualities or characteristics by which something is recognized) "it is the nature of fire to burn"; "the true nature of jealousy"

(n) nature (a causal agent creating and controlling things in the universe) "the laws of nature"; "nature has seen to it that men are stronger than women"

(n) nature (the natural physical world including plants and animals and landscapes etc.) "they tried to preserve nature as they found it"

(n) nature (the complex of emotional and intellectual attributes that determine a person's characteristic actions and reactions) "it is his nature to help others"

(n) nature (a particular type of thing) "problems of this type are very difficult to solve"; "he's interested in trains and things of that nature"; "matters of a personal nature"

'landscape' (L) glosses

(n) landscape (an expanse of scenery that can be seen in a single view)

(n) landscape (painting depicting an expanse of natural scenery)

(n) landscape, landscape painting (a genre of art dealing with the depiction of natural scenery)

(n) landscape (an extensive mental viewpoint) "the political landscape looks bleak without a change of administration"; "we changed the landscape for solving the problem of payroll inequity".

All indices of similarity and relatedness for the three groups of selected lexemes (i.e., direct event descriptor (F), candidates for semantic drift (RW) and *benchmark* words (NL)) are presented in Fig 3.2.

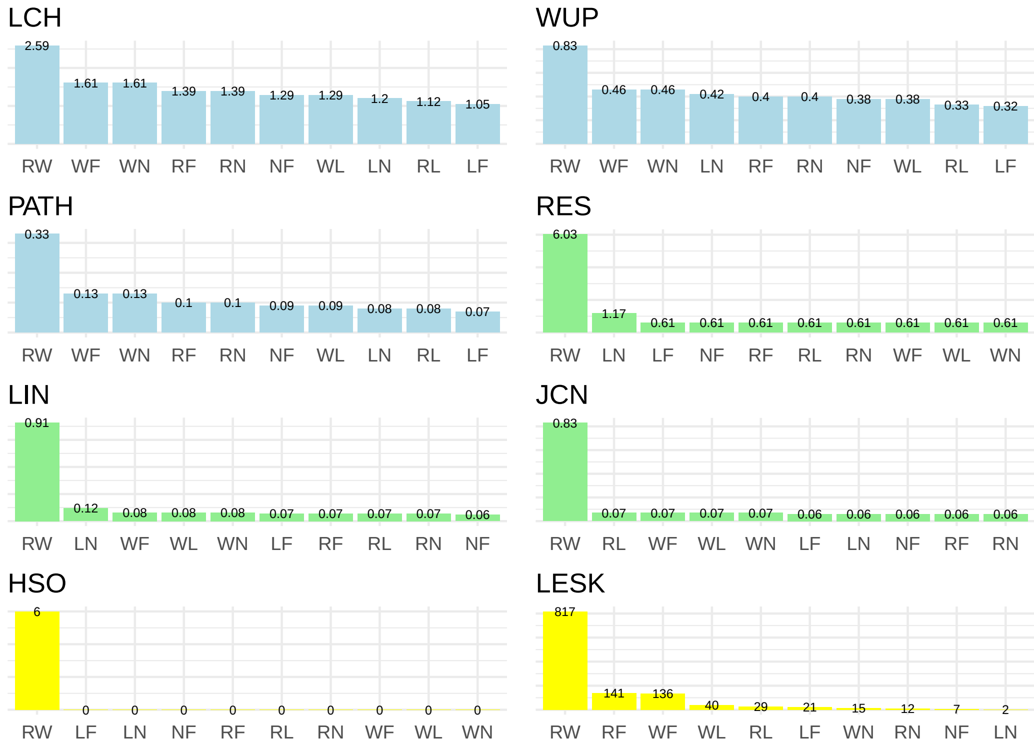

Figure 3.2: WordNet-based topological similarity (LCH, WUP, PATH, RES, LIN, JCN) and topical relatedness (HSO, LESK) measures between (FRWNL) lexemes, selected for the data experiments.

Knowledge about topological similarity and topical relatedness can be useful not only for the manual construction of the methodological frameworks; Given the growing recognition of the multidimensional properties of data, graph artificial networks are gaining importance [Scarselli et al., 2009] and therefore opportunities emerge to train new generation of Graph Neural Network (GNN) on the WordNet knowledge base [Hamaguchi et al., 2017; Saedi et al., 2018] in order to provide new event-specific semantic augmentation models (Fig 3.3).

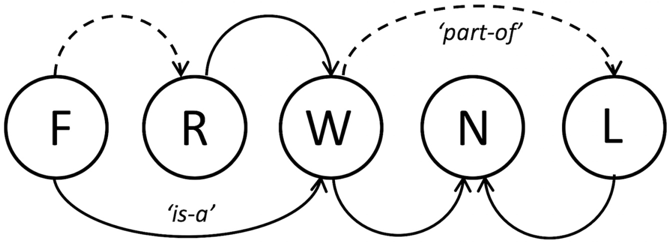

Figure 3.3: Graph-like representation of the relations between lexemes, constituting event-specific category *flood*.

3.3.2 Authoritative benchmark datasets

Three computational linguistic experiments, constructed to obtain empirical confirmation for several approaches towards re-purposing semantic drift concept for event-based data extraction are benchmarked in this thesis against three types of the flood monitoring information:

(i) past weather events,

(ii) geographically-approximated[1] locations of hydrometric (water levels) monitoring stations (e.g., river gauges and groundwater sensors), and

(iii) flood warning communication typologies (a.k.a., 'flood stages'), used to reflect the severity or potential danger levels of the approaching hazard, or to warn about status change of the ongoing flood event.

Past weather events

Met Office UK regularly issues and updates information regarding past weather events [PWE, 2019] (1990-), which cover floods, major storms, ex-hurricanes, hot spells, snow periods and low temperatures; The descriptive information is accompanied with the precise dates range.

Hydrometric monitoring points

Hydrometric monitoring points is the composite national (England and Wales) dataset consisting of four main components, notably, (i) surface water monitoring stations, (ii) river flow gauges, (iii) precipitation monitoring stations and (iv) groundwater levels observation stations. This dataset are available in GIS SHP format from the Government Open Data portal [DSP, 2019].

[1]Exact coordinates of each monitoring station has been anonymised for open data purposes by snapping each point to the nearest intersection of OSGB 1936 grid reference system, scaled down to 1km squares.

Flood stages and risk communication

Flood stage is used to describe the progress in covering the designated flood risk areas with water. It is defined by the NOAA National Weather Service as "an established gage height for a given location above which a rise in water surface level begins to create a hazard to lives, property, or commerce." [NWS, 2019]

The main principle behind the designation of flood risk areas is *topographic gradient*. Topographically dependent water movements define the convergence of streams, which are - if not intercepted or infiltrated - run from the higher locations to the lower ones, with the velocity, defined by the gradient of the routing slope [Tewolde and Smithers, 2006]. Derived initially from direct geodesic surveys, the most recent production of topographic maps involves applications of remote sensing techniques, predominantly radar and LiDAR based. The high resolution of LiDAR data also enables modeling of pluvial flood events by being able to capture the finest sinks and obstacles within impermeable urban structures [Diaz-Nieto et al., 2012]. Designation of topographically defined flood risk areas was also used for automatic classification of the properties contained within the designated boundaries as the ones being at risk. Depending on how flood stage progresses, flood risk areas are used by authoritative environmental bodies in order to make decisions about how to inform public and organize rescue and evacuation campaigns (Fig 3.4).

Figure 3.4: Three types of risk communication messages used by the Environment Agency in the UK: (a) Flood Alert ('Flooding is possible. Be prepared.') is used two hours to two days in advance of flooding; (b) Flood Warning ('Flooding is expected. Immediate action required.') is used half an hour to one day in advance of flooding; (c) Severe Flood Warning ('Severe flooding. Danger to life.') is used when flooding poses a significant threat to life.

Just like information about hydrometric monitoring points, datasets describing flood alert, risk and severe warning areas are available under the open data conditions from the Government Data Portal.

3.4 Methods

The choice of methods outlined below was mainly conditioned by the data modalities (i.e., lexical, temporal, spatial and visual), however, the common underlying theme was motivation to highlight *relational* processes both within and between modalities of the composite dataset, generated during particular event. I therefore used methods from statistics, geography, computational linguistics and computer vision, the disciplines, which are currently methodologically concerned with representations of interactivity between data structures [Xu et al., 2019].

3.4.1 Time series cross-correlations

One of the problems I started the experimental part of this thesis with and which is described in Chapter 4 is associated with finding relationships between several time series: direct event descriptors (x_t), candidate variable(s) (y_t) and benchmark lexemes (z_t). In situations like this not only we are interested in measuring the joint dynamics between the lexemes, we are also looking to find how series x_t may be related to past lags of the y-series, which means that the cross correlation function (CCF) can assist in identifying lags of the y-variable that might be useful predictor of x_t. Apart from solving the empirical problem of finding at which lag sample correlations between y_{t+h} and x_t (h = 0, ±1, ±2, ±3, etc.) are the strongest, I am also seeking to confirm or reject the hypothesis that crowd-generated production of the lexemes that are direct event descriptor candidates or benchmark material happen independently of each other. If shuffled correlations in their various configurations (for details, see Chapter 5, Methodology section) are statistically stronger than the continuous data streams then the process can indeed be seen as independent (e.g., if tags (L) and (R) co-occur in one message and (F) in the other during the same time interval, then tag (R) is not necessarily semantically closer to the neutral in the context of flood risk situation tag (L), it can be still related to the flood event).

3.4.2 Spatial autocorrelation

Studies that illustrate spatial linguistic interactions are quite scarce, since geostatistical studies are quite parochial and are much more likely to be seen in the domain of social geography (e.g., crime analytics) or ecology (e.g., dynamics of the species distribution), rather than in language studies. Nevertheless, it can be argued that lemmatised linguistic content is not much different from the ecological dynamics, especially when data is derived from social media and provides samples large enough to be able to be reliable (however, also arguably largely corpus-specific) insights into the ways people communicate about certain phenomena in different areas. There is therefore an opportunity to reveal how spatial interaction of our flood-related lexemes-tags can lead to semantic drift, which increases the volume of the event-relevant data in a statistically significant way - which I follow up upon in Chapter 5.

Conventional statistics are largely based on the assumption that variables are *random*, *independent* and *identically distributed* ('i.i.d assumption') [Durbin, 1973; Gaenssler and Stute, 1979], however, this assumption can only be considered reasonable when samples are taken from controlled experiments where variables are assumed to not interact with each other, which is very a rarely realistic scenario. In many cases, the *i.i.d. assumption* is convenient because it allows ignoring the entire domain of complex relationships random variables may form with each other when drawing inference. However, if not taken into account, this can cause an overestimation of important underlying patterns, as well as an obscuration of characteristics of individual random variables when samples are not identically distributed [Westerholt et al., 2018].

Social media derived data that has a spatial attribute (or modality as defined in the context of this thesis) is not collected in a fully controlled manner but taken from in situ physical and social contextual conditions [Goodchild, 2009]. One of the applications it being used for is different socio-geodemographic mosaics (e.g., Experian Mosaics), where data is often aggregated into arbitrary units that are defined for purposes other than mere spatial analysis, e.g., census statistics. In such conditions, it would be fair to assume that those adjacent spatial units may then be subject to similar contextual influences that are produced by the phenomenon of interest, the boundaries of which are much larger than those ones of arbitrarily defined spatial units (e.g., socio-political unrest or species distribution). These characteristics inevitably cause redundancy within related random variables, which, in turn, violates the *i.i.d. assumption*. These circumstances motivate the first Law of Geography, which states that "everything is related to everything else, but near things are more related than distant things" [Tobler, 1970]. This law was also statistically confirmed by the concept of spatial autocorrelation, which is a second-order characteristic that describes the spatial interaction behaviour within random variables [Fischer and Getis, 2010]. A range of different estimators exist [Getis, 2007, 2008]: the covariance-based Moran's I and Geary's c [Cliff and Ord, 1969], the spatial autoregressive coefficients ρ (autoregressive parameter for the spatial lag model) and λ (autoregressive parameter for the error lag model) [Anselin et al., 2008], or G_i [Getis and Ord, 1992], which emphasizes structures among extreme values. These estimators evaluate spatial interaction behaviours within random variables but are used in different application scenarios. These include the assessment of influences of distance effects, of the roles of *geometry* and *topology*, or of the impact that individual geographic features have on spatial processes [Getis, 2007].

In this thesis, I am particularly interested in the multivariate configuration of Moran's I [Moran, 1950], which is one of the most popular estimators of spatial autocorrelation. Simple Moran's I measures the normalized spatially-weighted covariance within random variables, where it takes account of geographic structures by incorporating a so-called *spatial weights matrix*. Spatial weights define fixed geographic structure connecting those spatial units $s_i \in S$ upon which the investigated phenomenon is believed to operate [Bavaud, 1998; Harris, 2011], however, many geostatistical packages (e.g., GeoDa) offer automated quantification of the weights' matrices by running preliminary iterations on the distributions of the data points. Using Moran's I as a test statistic allows the investigation of whether the modelled geographic layout plays a significant role in the structure of an attribute, and thus if geographic factors can be regarded as major drivers of interactions within the analysed random variables.

The multivariate configuration of Moran's I used in this thesis is in principle very similar to the Moran's I tool, however, rather than determining the level of spatial autocorrelation within one variable (that is, how clustered in space one variable is in terms of high and low values), it determines whether there is spatial autocorrelation between two or more variables. Like the Moran's I, the range of possible values lies between -1 and 1: An estimate of 0 implies no spatial autocorrelation, more significant estimates move closer to 1, the greater the degree of positive spatial autocorrelation while the closer they get to -1, it indicates stronger negative spatial autocorrelation. This method was the principal statistical instrument for answering my second research question defined in the Introduction and detailed methodology for which is described in Chapter 5.

3.4.3 Lexical network analysis

Chapter 6 of this thesis represents the final experimental part of my research and aims to answer how meaning dynamics is reflected in the visual modality of social media posts created during various stages of flood events. The main two sets of methods used here are *lexical network analysis* and *word embeddings*.

Since Lakoff's work on semantics, it has become a commonplace practice within the framework of cognitive linguistics to organize the different senses in a form of a lexical network, where prototypical sense in placed in the centre and increasingly peripheral senses are moved away from it [Brugman and Lakoff, 1988]. As a matter of fact, such networks were established with no prior empirical evidence, relying on what Sandra and Rice (1995) refer to as "unshakeable confidence in introspections of linguists". This approach has been criticized for its arbitrariness by corpus linguists claiming that any lexical network requires empirical validation [Gilquin, 2008].

As has been already mentioned in the literature review, categories are constructed under a number of necessary and sufficient conditions, the principal ones of which are:

(a) any element should satisfy these conditions if its considered to be a part of that category,

(b) consequently, some members are more representative of the category than the others, and

(c) consequently again, some senses of a polysemous word become so typical, central that they turn into categorical *prototypes*.

The shape of such structures is known as a 'radial network' [Lakoff, 1987], however, I am personally not happy with such definition as it assumes the independence of each axis and hence excludes the possibility of horizontal synonymous relationships between the senses, which is not true, as Lakoff himself acknowledges the existence of the *lateral* (also known as 'synonymical') relationships between the senses [Brugman and Lakoff, 1988]. I therefore try to avoid in the context of this thesis the term 'radial network' and stick to the 'lexical network' concept.

In the majority of the cases, lexical networks are established on the basis of *intuition*, where the central sense of the word is being arbitrarily selected first, with the subsequent addition of the [seemingly] related lexical categories. On this account, [Sandra and Rice, 1988] pointed out that "there is a lot of vagueness regarding the nature of the represented reality, at both the linguistic and cognitive levels" as a consequence of such methodological design process. As a response to this criticism, both cognitive and linguistic validity have been tested in subsequent studies. The linguistic one was tested by means of frequency as attested in corpora, while cognitive validity was tested by means of a sentence assembling experiment designed to elicit the most salient sense of the word represented in the lexical network - i.e., by means of contextual validation.

While the problem of lexemic frequency has already been methodologically addressed and discussed earlier in this Chapter, the contextual validation remains a challenge since we are dealing with *de-contextualised lexical material* (i.e., tags). In order to address this cognitive gap, I turned to the accompanying visual modality in order to identify mechanisms behind linguistic construals. If we associate each tag with the image and take a look at its components, we can quickly identify semantic links between the object in focus of the image and its linguistic annotation. Moreover, it can be claimed that other objects that maybe more peripheral or less well distinguishable visually may be also semantically linked to the focal elements and these relationships are based on contextual (or event-driven) semantics, rather than on the dictionary one, so the image structure quickly acquires the properties of a network analogous to the linguistic one described above and based on the same-sense principles. And if we follow my earlier comment, it's not a *radial network*, but rather a *complete undirected weighted graph*, where all the objects are connected and the strengths of those connections is defined by their cosine similarity in the event-driven (domain-specific) corpora. However, since the spatial scale of the process of interest is rather large, instead of looking at the structure of every individual image, I look at the ensemble of images tagged with the same linguistic tags, where each image is classified as a scene by a deep learning Convolutional Neural Network (CNN). I propose an assumption that people tend to focus their attention on similarly structured scenes during the various stages of the evolving events and this property can be used to capture various stages of, in our instance, flood events when analyzed as aggregated ensembles. As metrics, I use the notion of attention focus, since it has been recently established that the natural scene perception requires attention [Cohen et al., 2011]. From the behavioural geography perspective, attention focus tends to coincide with the *objects* that belong to the category of 'landmarks': They are used to denote familiar objects or scenes that can aid environmental navigation. If the landscape is familiar, then it consists of a high number of recognizable landmarks with a high degree of semantic similarity between them, and this type of navigation is referred to as 'route following' one. In contrast, compilation of poorly identifiable, disconnected scenes is attributable to situations as seen by people unfamiliar with the area, and hence they exhibit certain characteristics of 'wayfinding' behaviour.

Attention focus metric is derived from the probability of CNN scene classifications and *attention density* is derived from the semantic network density value. These values are subsequently used to derive information about the spatial mobility of flood event participants around the timing when flood risk communications of various degrees of severity have been made by the Environment Agency during the period 2004-2014. More detailed description of the method is presented in Chapter 6.

3.4.4 Word embeddings

In order to extract values of the density of semantic networks for the previous method that have at their nodes CNN classified images-scenes tagged by any of the three groups of our interest (risk-signalling (or negative) (F); positive (NL) and neutral (RW)), I used values of cosine similarity metrics of word embeddings to construct the edges and estimate network densities.

Computational modelling of linguistic meanings is based on the assumption that they can be inferred from their immediate syntactic context (or 'word embeddings'). According to [Rohrdantz et al., 2012], research in this area mainly focuses on two objectives: word sense disambiguation (WSD) and word sense induction (WSI). The goal of WSD is to classify occurrences of polysemous words according to manually predefined senses, whilst the aim of WSI is to learn word senses from text corpora without having a predefined number of senses - where the later goal is much more harder to achieve.

Since one of the research interests of this thesis is linked to the WSD task, it is worth mentioning that the two popular methods for performing such a classification are word2vec and Latent Semantic Analysis (LSA) [Deerwester et al., 1990]. [Sagi et al., 2009] also have demonstrated that broadening and narrowing of word senses can be tracked over time by applying both methods to 'small world' contexts in diachronic corpora and conclude that word2vec captures similarity in a better manner. Some other algorithmic implementations of word embeddings also exist, notably skip-grams with negative sampling (SGNS), Global Vectors for word representations (GloVe) and Positive Pointwise Mutual Information (PPMI) [Navigli, 2009].

Word embeddings in general are a popular technique in natural language processing (NLP), often regarded as shallow technique owing to its computational efficiency, where words from the dictionary are mapped to low-dimensional vectors. These models can be either easily trained by the user or are publicly available via several implementations from the web. In this thesis I used the pre-trained *Google word2vec model* that was trained on Google News data (appr. 100 billion words) and contains 3 million words and phrases fitted using 300-dimensional word vectors. Their increasing popularity, as compared to more traditional approaches in computational linguistics, such as distributional semantics models, is motivated by the fact that they seem to continuously and significantly outperform other methods, supposedly due to their neural architecture, which allows prediction of words, rather than simply counting their co-occurrences [Levy et al., 2015]. As a consequence, embeddings are increasingly being used by researchers in novel and creative ways, especially in fields such as digital humanities and computational social science [Hamilton et al., 2016a; Bakarov, 2018], where dependency parsing, named entity recognition and bilingual lexicon induction are just a few examples where the use of embeddings as features has increased performance in recent years.

In technical terms, word embeddings are mappings of words to points in a K-dimensional continuous space, where K is much smaller than the size of the vocabulary. This dimensionality reduction has two main advantages: firstly, large and sparse vectors are transformed into much smaller and denser vectors; and secondly, the conflation of features uncovers latent semantic relationships between words. These semantic relationships are usually measured via cosine similarity, though other metrics such as Euclidean distance and the Dice coefficient are also widely used.

In NLP, word embeddings are often used as features for downstream tasks, specifically in studies concerning contemporary issues in language and culture. For example, [Hamilton et al., 2016b] trained separate embeddings on temporal segments of a corpus to track changes in the similarity of words to measure semantic drifts and [Heuser, 2016] used embeddings to characterize discourse about virtues in 18th Century English text. Other studies used cosine similarities between embeddings to measure the variation of language across geographical areas [Kulkarni et al., 2015] and time [Kim et al., 2014], and it is worth mentioning that each of these works sought to "reconstruct the mental model of authors based on documents", an application that will be exploited also in this thesis when attempting to answer my third research question.

There essentially exist two NLP perspectives on the usage of word embeddings, notably downstream- and corpus-centered ones, which differ by their approach to exploration of implicit bias in word embeddings. From a downstream-centered perspective, these stereotypical associations representing bias should be filtered out before using the embeddings as features; and in contrast, from a corpus-centered perspective, implicit bias in embeddings is not a problem that must be fixed but rather a means of measurement, providing quantitative evidence of bias in the training corpus.

It should also be remembered that embeddings are not a single objective view of a corpus - and much less an objective view of language. The corpus is itself only a sample used to represent the domain or phenomena of interest to the researcher and the ways this sample is being curated (e.g., modification of its size, length of the documents or inclusion of specific documents) can cause significant variability in the embeddings. Also, different algorithms demonstrate different corpus sensitivities.

For example, LSA, GloVe, SGNS, and PPMI are not sensitive to document order, however, all four algorithms can be sensitive to the presence of specific documents (though this effect has been proven weaker for PPMI case) [Antoniak and Mimno, 2018]. Some additional studies have demonstrated that the use of embeddings as sources of evidence of semantic proximity between meanings needs to be tempered with understanding that cosine similarities are not always reliable metrics and that smaller corpora and longer documents are more susceptible to variation in cosine similarities between embeddings [Bakarov, 2018; Antoniak and Mimno, 2018].

In my third data experiment, described in Chapter 6, I used the largest pre-trained vectors trained on part of Google News dataset [w2v, 2013] containing about 100 billion words and the maximal length of the documents are two words, corresponding to description of the natural scenes from the MIT Places database [MIT, 2014]. These properties of the data selected for my empirical analysis are expected to partially mitigate known issues associated with application of cosine similarity metrics.

Chapter 4

Event detection with semantic microchange on social media

4.1 Synthesis

As useful as word embeddings may be for answering questions about how language works diachronically, the non-discriminatory approach they use for the contextualization of words is currently poorly adapted for event monitoring. Although [Hamilton et al., 2016b]'s embeddings quantify semantic change for thousands of words and have proven to be a robust and valuable tool for complex overarching contexts, some studies (e.g., [Antoniak and Mimno, 2018]) have pointed out some risks associated specifically in regards to *under-representation* of the *fine temporal variability* within corpora and sensitivity to the presence of specific documents [words], which can been seen as a problem in event analytics.

Secondly, the language of events is also *affective*, and although there are much more negative words in the dictionary, positive words tend to be more frequent [Iliev et al., 2016]. Following [Hamilton et al., 2016a]'s work on *sentiment drift*, the implications are that negative words are also more diachronically unstable, i.e., they have faster rates of semantic change. Since this is contradictory to my target of *maximisation of the event-driven unstable lexemes*, here I examine how lexemes-candidates (RW) with *close-to-neutral* sentiment scores (0.0; -0.006, respectively) co-evolve with the direct event descriptor (F |-0.03), whilst being benchmarked against more positive concepts of (L |-0.02) and (N |0.25). My results indicate that event-specific neutral words do tend to change their meaning from positive to negative around flood events.

4.2 Background

Contemporary environmental hazard forecast mechanisms are based on highly specialised procedures, often involving cross-institutional partnerships, like, for example, the Flood Forecasting Centre [FCC, 2019] in the UK, created in 2009 as a partnership between the Environment Agency and Met Office. What they offer is a range of services, packaged as *Flood Guidance Statements* (FGS), which aim to provide information to citizens to help them with emergency planning and decision-making when facing or experiencing any form of natural flooding, including from river, surface water, tidal/coastal or from groundwater sources. Fairly recently (2016-2017, exact date unknown) they have extended their 2-day advance regional warning mechanisms towards 5-day ones [FWI, 2019] for England and Wales, with detailed reference to background weather information [MOF, 2019] aimed at explaining mechanisms behind approaching or evolving natural hazards.

The models behind flood warnings systems are gradually evolving towards much *finer resolutions* (e.g., city scales) and much more *hybrid* (e.g., designed for both atmospheric research and operational forecasting applications) tools. Examples of such systems are *Weather Research and Forecasting Model* (WRF) [MMM, 2019] in the US and the *Unified Model* (UM) [UM, 2019] in the UK. The first one was designed to serve a wide range of meteorological applications across a range of scales (from tens of metres to thousands of kilometres), and has been particularly successful at city-sized mesoscale simulations at the sub-kilometre scale, whilst the UM model has only gone as far as regional scale, failing to break through the sub-kilometre threshold and thus fuelling a growing desire in the research community to achieve this standard in seamless UK meteorological prediction systems. For example, one research project conducted at the University of Warwick (UK) aimed to provide a set of reliable, high-resolution reference simulations over a range of environmental conditions in order to understand why UM calculations do not agree with ground observations made by radar [Sit, 2017]. This illustrates that flood forecasting and monitoring are being increasingly characterised as a 'big data' problem, where a lot of hope is pinned on using different data sources, such as satellites, radar systems, rainfall gauges and hydrological networks, integrated and processed with help of the parallel computing and other state of the art infrastructures [Degrossi et al., 2017].

Apart from the geo- and atmospheric sciences research community, this interest in finer-scale predictions has also been picked up by interdisciplinary researchers, looking at how, for example, 'social sensors' (i.e., multimodal signals generated on the web and social media) can fill in gaps in the demand for sub-kilometre resolutions [Stopczynski et al., 2014]. Of particular interest are social media messages with geographic reference, also known by the term of *volunteered geographic information* (VGI) [Goodchild, 2009] as they can provide some highly valuable insights of what is happening in a specific location with help of text and images often provided alongside temporal and geographical attributes of postings. However, due to the sheer volume of these uploads and their semi-structured nature, one of the outstanding challenges remains to understand how to deal with such data signals, i.e., how to 'separate wheat from chaff' without losing the volume and resolution of useful information. This conversational paradigm can also be seen as an extension of other, more analogue, but equally unstructured real-world behaviours reflecting human action facing environmental uncertainties [Tkachenko et al., 2015].

Information seeking and sharing, and implementation of resilience measures can be regarded as a particular type of risk-signalling behaviour that is absent from the contemporary warning systems [Tkachenko et al., 2017c]. Tracking the ways in which people not only behave but also make changes to their properties and environment can open a new direction for behavioural research, specifically, for environmental habits tracking and re-calculation of risks. Often seen as a reactive response to external stimuli, social media postings have demonstrated some potential for estimating socio-economic impacts of natural hazards, where degree of success is defined by the recurring nature of the event, its spatial footprint and local enthusiasm and proactive behaviours for researching information about emerging flood risks [Tkachenko et al., 2016a, 2017b].

Nevertheless, this proliferation of social media platforms has introduced a new and additional source of information to be taken into account when designing *warning systems* and planning their implementation. Thus, the US Geological Survey (USGS) was the first environmental institution to recognise the value of such user generated content (UGC), acknowledging that analysis of the content and geographic distribution of Twitter postings - i.e., 'social sensors' - can be a useful supplement to instrument-based estimates from physical sensors of earthquake location and magnitude [Earle, 2010]. A more recent study reported on the value of Flickr image sharing platform in *nowcasting* of the evolving ex-hurricane Sandy, where aggregated volumes of UGC was found to replicate air depression fluctuations over the same time period [Preis et al., 2013]. In these early studies, various open UGC is treated as a valuable nowcasting tool in conditions where official observation stations, sensors or gauges do not provide sufficient geographic coverage to be able to generate timely and spatially accurate updates about the situation on the ground.

Despite growing interest in these new and widely available data sources, current methods of using social media data have begun to reach the limits of their potential. We argue that the explanation lies somewhere in *how* 'useful' data components are being defined. For example, all current analyses are based on the exact words and word-combinations designating either type of a hazard itself (e.g., 'flood', 'hurricane') or its name (e.g., 'Sandy', 'Katrina') [Preis et al., 2013; Hilfinger et al., 2011]. However useful for operational purposes, from the *forecasting* perspective they hold relatively little value as they limit these initiatives to their mere nowcasting capacity.

Turning attention to the origins of the reasons behind people's conversations about the natural world both in analogue and digital worlds, it was decided to seek answers in the area of environmental semantics, which defines lexical structures people use in order to express their interactions with the natural environment (see Chapters 2 and 3). This decision was linked to an initial hypothesis that any natural phenomenon can be described by a much wider spectrum of words and structures than those that are currently being employed by social media analysts. For instance, environmental anthropologists recognise that people's engagement with, for example, water in the landscape - or the landscape itself - can be seen as a cognitive experience of endlessly changing states [Tkachenko et al., 2017c]. Each of these states has its own qualities and associated meanings, which are always potentially there, detectable by human sensory experiences in many different contexts and under various circumstances [Wohlleben, 2018]. However, to the best of my knowledge, no study has been conducted to exploit this research opportunity in order to verify the value of information encoded in these accompanying meanings, i.e., where they stand as compared to the direct event descriptors.

In this data experiment I therefore aimed to test the novel hypothesis that alternative environmental semantics (or speaking methodologically, *lexemes-candidates for semantic drift*) can be linked to extreme environmental events and may serve as a predictor of an evolving hazard.

4.3 Hypothesis

In the scope of this analysis I propose a hypothesis that in the context of the negative hazard event (i.e., flood), topical neutral words are more semantically unstable than emotionally charged ones - and therefore will have tendency to fluctuate towards negatively-charged direct event descriptors and away from more positive environmental descriptors.

4.4 Methodology

4.4.1 Variables

In many respects, current approaches to the selection of words for event monitoring can be seen as somehow naïve, compiled with help of the crowd-sourcing (e.g., via former *CrowdFlower* (now *Figure Eight*) platform) or selected arbitrarily by researchers ('armchair linguistics'), where both are largely intuitive, random or entirely usage-driven. Such *armchair* approaches contrast with the *corpus methods*, where words-candidates are selected by algorithm and without much account of human perception, attitude to the phenomena in question or their subjective preference of one word over the other. I argue that the truth lies somewhere in between both approaches. On the one hand, we cannot discard the objective computational part, however, we still need to make sure that the subjective component is accounted for. Therefore, in the scope of this work, requirements for data entries (variables) were established as follows:

(i) Word-candidates for event-predicting semantic drift should be able to demonstrate some general positive trends of temporal co-evolution with the direct event descriptors;

(ii) They should have sentiment tone that is slightly different, preferably more neutral than the sentiment of direct-event descriptors, which are expected to be more emotionally charged depending on the nature of the event (positive or negative);

(iii) They should also be able to demonstrate a positive correlation with some other background lexemes, which, in turn, should not be strongly co-related to the direct event descriptors.

The idea of selecting variables on the basis of their *correlation* (i.e., 'proportional co-occurrence') capacity with direct event descriptor(s) is linked to the hypothesis that such an overall trend is composed of periods of very strong and weak correlation, where the latter should logically coincide with elevated rates of co-occurrence with background lexemes. Regarding the sentiment component of requirements, the advantage of using the [near]neutral words lies in their expected predictive capacity. Their neutrality assumes their penetrative usage outside event boundaries, which are socially constructed on social media with help of direct event descriptors. Hence, we can hypothesize that they can be useful for either event prediction, its impact estimation - or both. In previous preliminary studies I tried to analyze how good digital traces from several platforms (Google Analytics, Flickr) are at predicting or estimating *impacts* of several natural hazards, including floods.

Interest has therefore evolved towards understanding the behaviour of neutral words around *emotionally charged* events.

4.4.2 Cascading non-probabilistic sampling

From the method design perspective, according to [Antoniak and Mimno, 2018], who evaluated the stability of embedding-based word similarities, it has been found that relatively reliable results can be obtained by simply averaging over multiple bootstrap samples in all tested cases. I therefore decided to use the non-probabilistic sampling design in order to extract lexemic behaviour to compare longitudinally.

Non-probability sampling is generally known as a sampling technique where samples are gathered in a process that does not give all the individuals in the population equal chances of being selected [Higginbottom, 2004]. As compared to probability sampling, non-probability sample does not result from a randomized selection process. Here, variables are usually selected on the basis of their accessibility or by the purposive, personal judgement of the researcher. Whilst this method is popular in a number of instances (outline below), some of the obvious downsides are that it is not always clear what proportion of the entire population is being represented, often leading to conclusions that the results are not generalisable.

Non-probability sampling can be fit for purpose in the following situations:

(i) When there is a need to highlight a particular trait in the population,

(ii) When the aim is to perform a qualitative or exploratory study, or

(iii) In pilot studies, which will be subsequently followed by experiments utilizing randomized probability sampling.

Within the scope of this data experiment I selected variables (lexemes) associated with crowd postings on the photo-sharing platform Yahoo! Flickr (i.e., 'flood' (F), 'river; water' (RW), 'nature; landscape' (NL)), with the specific purpose of extracting their cross-correlations. This approach is justified by 'armchair linguistics' [Fillmore, 1991], where the researcher is allowed to make some arbitrary data selections if they believe that some subjects are more fit for purpose as compared to other individuals.

Method development began with the novel hypothesis that some alternative tags, which satisfy the criteria outlined above, can exhibit event-detecting behaviour in the context of emerging flooding. This assumption was motivated by the ways people tend to interact with their surrounding environment, which changes as precipitation builds up or water levels start approaching residential areas (or areas that are not normally covered with water). Here, I argue that during a certain time period before the peak of the event, the semantics of human conversations on social media also become distinct from the general context meaning [Spirn, 1998] and acquires new properties under the influence of the new perceptual experience, such as, for instance, observations of saturated soil or raised water levels in nearby streams [Strang, 2004]. As the primary aim of this study is to verify the predictive potential of social media platforms on their own (e.g., without combining them with data signals from other sources, such as river gauges or precipitation readings), I have therefore defined *events* in my dataset as 'flood peaks' on social media, reflected by maximum uploads of content tagged with the direct event descriptor (i.e., (F)), around the time of flood events, identified by several authoritative sources (Fig 4.1). Since this was an approximate selection, scope still remains for future data experiments to define social media indicators suitable for cases of *unpredictable events*, such as various natural hazards since similar sets of indicators have been already developed for social events. The latter ones are regarded as more *predictable* as they are usually scheduled in advance, their time of occurrence is usually known and hence there is 'before' and 'after' data from social media platforms available for comparison. In contrast, for *unpredictable* events we can only get data after the event [Fujiyama et al., 2016].

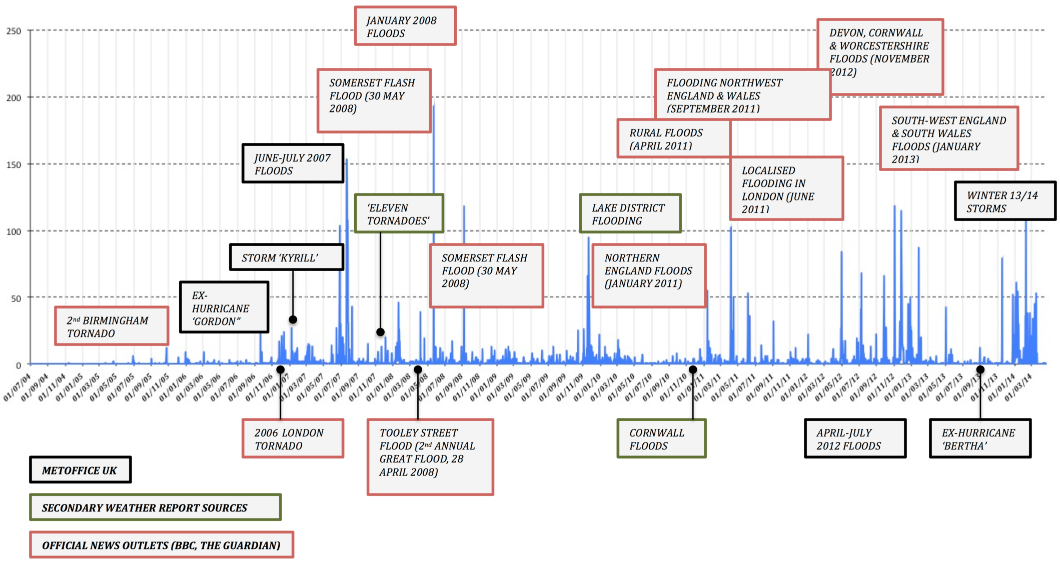

Figure 4.1: Timeline of Flickr activity during a number of flood events(2004-2014) of various magnitude, captured by official, secondary and massmedia sources.

For the algorithm development, I needed three conditions to be satisfied in order to prove the predictive capacity of lexemes-predictors:

(i) Be responsive to changes in event magnitude, which is represented by the number of daily uploads of the direct event descriptors to the social media;

(ii) Their behavioural profiles should differ within and outside designated 'event buffers': For example, during the time period around the event outbreak they should exhibit statistically significant correlative behaviour with the direct event descriptors, whilst outside those periods these relationships change and they may potentially start correlating with other sets of topics; and

(iii) Lexemes-candidates have to exhibit stronger correlations with direct event descriptors in any period preceding the 'flood peak' time interval, where 'flood peak' represents socially-constructed start of the event (according to the known log-normal distribution of the *unpredictable events* [Preis et al., 2013; Fujiyama et al., 2016] on social media). In this respect, it can be argued that this type of flood prediction is a real case of event detection as we need to capture *socially significant* natural events, rather than hydrological phenomenon, which do not always involve people or imply negative consequences for communities.

The adaptation of these conditions to algorithm construction for the purpose of this study is as follows: I select three sets of lexemes: **(1)** direct event descriptor(s), which are usually hazard-naming words, such as 'flood', 'earthquake', 'storm' etc., **(2)** benchmark words, which have a more general meaning, exhibit positive sentiment and demonstrate correlation trends with the words-candidates, but not with the direct event descriptors, and **(3)** words-candidates for event detection, which possess intermediate sentiment value between direct event descriptors and benchmark lexemes, and exhibit correlation trends with both sets of words.

Here algorithm development began with four selected input tags ((N), (L), (R), (W)), plus two additional aggregated tags ((RW) and (NL)). Each of these inputs was connected to a single output (combined risk-signaling tags 'flood', 'flooding', 'floodplain'), without any initial weight attribution. As an activation function a simple Pearson Product-Moment Correlation was used, which is an advantage for our case study where linear output units are required:

$$\rho_{X,Y} = (\frac{cov(X,Y)}{\sigma_X \sigma_Y}) \tag{4.1}$$

The very first iteration (a) of the algorithm, which is presented in Fig 4.2 is therefore designed to verify whether selected words in our corpus tend to correlate in accordance with our initial requirement: $\rho_{(F)(NL)} \ll \rho_{(F)(RW)} \wedge \rho_{(RW)(NL)}$.

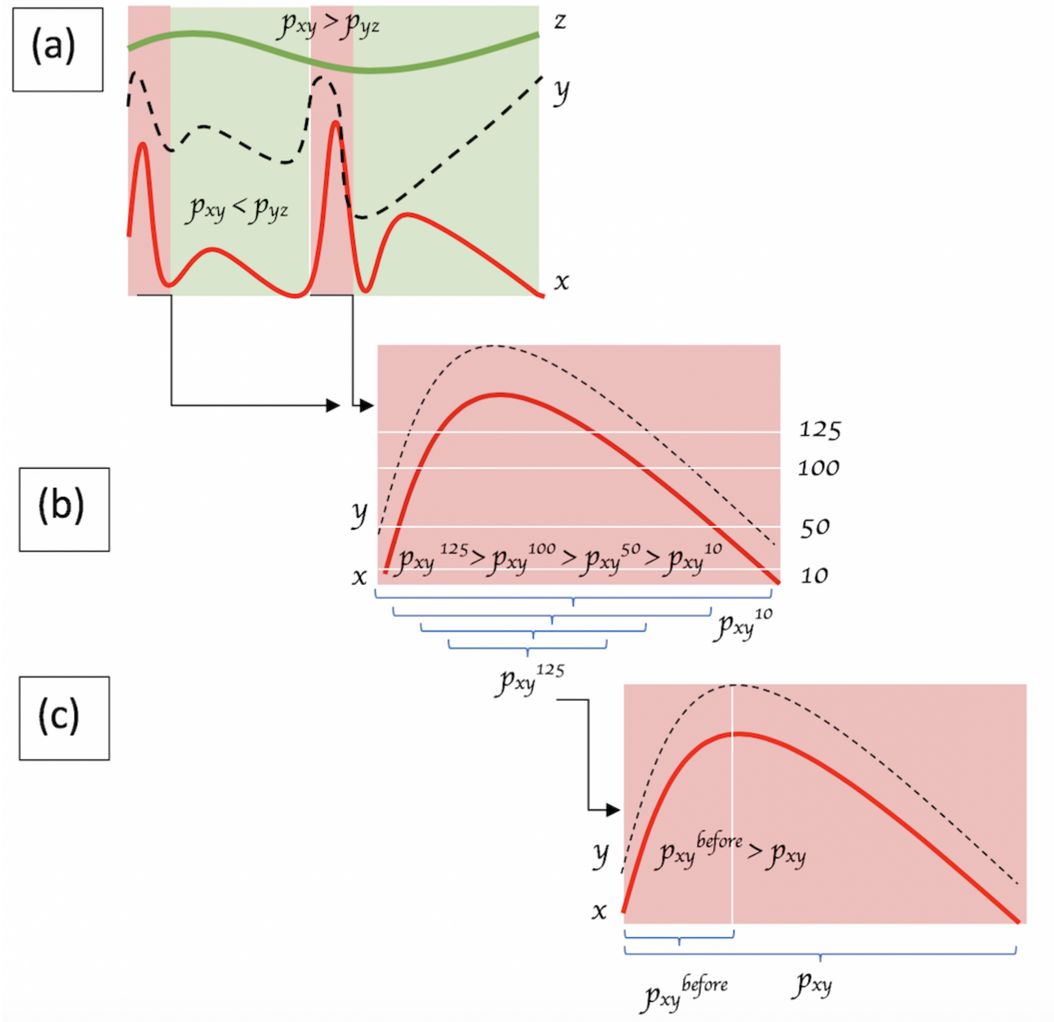

Figure 4.2: Conceptual workflow: **(a)** segmentation of the study period into 'flood-peak' and 'hazard-free' intervals; **(b)** definition of the max event severity with help of the daily social media uploads (here: 10, 50, 100 and 125 uploads per day); **(c)** segmentation of the flood-peak periods into pre- and post-climax periods, using max social media activity as a threshold.

Having satisfied the first set of requirements, I then looked at whether it was possible to capture the dynamics of those initially defined relationships between sets of lexemes, or whether these relationships are static. For this purpose a 5-day buffer was designated around each 'flood peak' (however, equation 4.2 illustrates that any buffer can be selected as it strongly depends on frequency of hazard events in particular locations), in order to make this method comparable to the current warning timescale by the FFC UK [FWI, 2019]. For the sake of this experiment, it was assumed that 'flood peak' on social media corresponds to the very beginning of the event, spiking at the outbreak and then again at each situational change according to the crowd interest. Therefore, each 5-day period preceding each event outbreak can be considered as 'forecast period', and each 10-day period around each peak is regarded as 'event buffer'.

$$B_{peaks} = (\frac{N_{all} - N_{peaks}}{N_{peaks}})^{\frac{1}{2}}, \tag{4.2}$$

where B_{peaks} is a number of forecast days, N_{all} is the number of days in the study period, N_{peaks} is a number of days with event-specific social media activity in the study period.

I then looked at the dynamics of lexemic co-occurrences inside and outside each 'event buffer', following corresponding set of conditions: **(a)** relationships between F and RW should become stronger inside the buffer across the dataset: $\rho_{(F_{inside})(RW_{inside})} \gg \rho_{(F_{outside})(RW_{outside})}$; **(b)** relationships between NL and RW should be stronger outside the buffers than inside: $\rho_{(NL_{outside})(RW_{outside})} \gg \rho_{(NL_{inside})(RW_{inside})}$; **(c)** relationships between NL and F should become even weaker inside buffer than they were continuously across the corpus timespan: $\rho_{(F_{inside})(NL_{inside})} \ll \rho_{(F)(NL)}$.

So far we have looked at the strength of the co-occurrences of our variables, without giving too much attention to the magnitude of the event, i.e., changes in daily post volumes, which could be indicative of dramatic turns during the course of the event. By introducing the magnitude condition, I assumes that the impact of the event is reflected in the number of crowd-generated uploads of alternative lexemes per day (e.g., more than 10, 50, 100 or 125 posting thresholds, to illustrate their sensitivity to the event dynamics/evolution), so the ideal would be a scenario where the event magnitude shows the strongest correlation between the direct event descriptor and the lexemes candidates inside the 'event buffer', which should visibly decrease when comparing days with a smaller number of uploads: $\rho_{(F_{inside125})(RW_{inside\cong})} > \rho_{(F_{inside100})(RW_{inside\cong})} > \rho_{(F_{inside50})(RW_{inside\cong})} > \rho_{(F_{inside10})(RW_{inside\cong})}$. Also, hypothetically, no difference should be evident when looking at the same dynamics outside 'event buffers': $\rho_{(NL_{outside125})(RW_{outside\cong})} \sim \rho_{(NL_{outside100})(RW_{outside\cong})} \sim \rho_{(NL_{outside50})(RW_{outside\cong})} \sim \rho_{(NL_{outside10})(RW_{outside\cong})}$.

Finally, we shift our attention inside to the 'event buffers', which are composed of ±5 days before the event outbreak (first local maxima of uploads for each event, mentioning the direct event descriptor), in order to find out whether our alternative tags could be used for event detection purposes: $\rho_{(F_{inside(5...1)})(RW_{inside(5...1)})} > \rho_{(F_{inside})(RW_{inside})}$, where n is a 'flood peak'.

4.5 Results

Fig 4.3 illustrates the dependency behaviours of tags inside and outside flood peak periods, without taking into account the magnitude of the event.

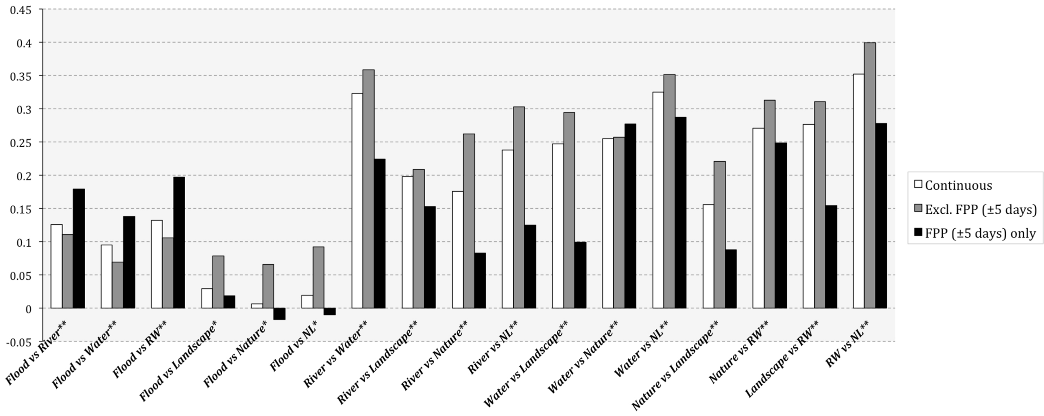

Figure 4.3: Correlation values between positive (NL), negative (F) and neutral (RW) tags on the *Yahoo! Flickr* platform (2004 - 2014).

These results illustrate that flood-related tags (F) tend to correlate with hydrologically themed tags (lexemes-candidates for event detection), which constitute the linguistic component of the uploaded content, either on their own (e.g., (R) or (W)) or in combination (RW). These dependencies illustrate the general contrast with how other tags, which are linked to more generic environmental thematics ((N), (L)), relate to the event detecting lexemes, which successfully satisfies our initial requirement for the input variables. It is also possible to confirm that (R) and (W) tags occupy an intermediate position between the topic of the natural hazard and more generic natural landscape theme.

Since combined RW tags demonstrated higher correlation rates with F tags inside flood peak periods and slightly decreased ones outside, Fig 4.4 shows dependencies between combined tags (RW) and flood-related tags outside (1a) and inside (2a) flood peak periods respectively, this time accounting for the event magnitude condition, where impact is reflected by the number of crowd-generated content uploaded per day (e.g., more than 10, 50, 100 or 125 photos).

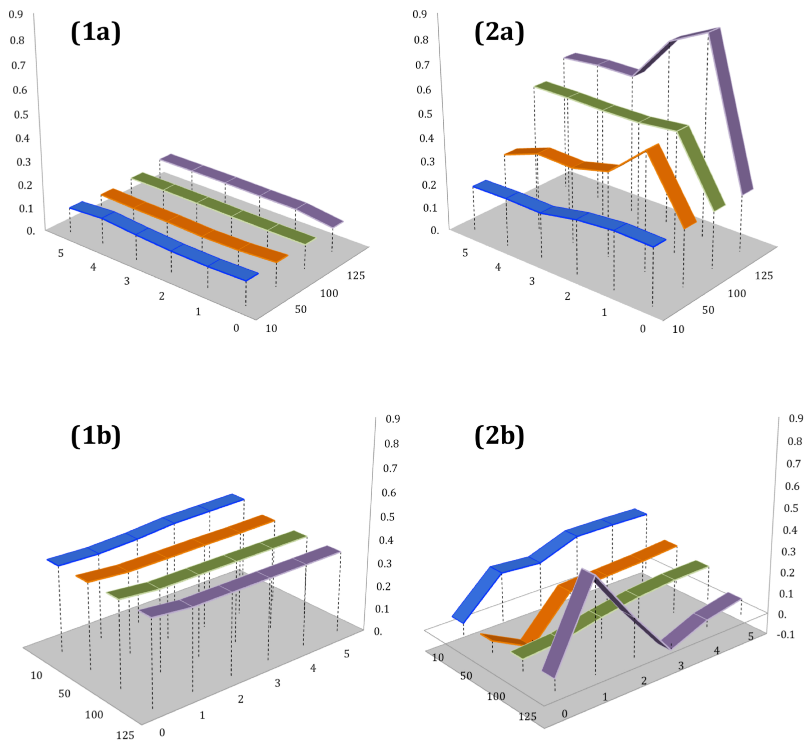

Figure 4.4: PPMCC values of the cross-dependencies between aggregated tag material (a) (RW) vs. (F) and (b) (RW) vs. (NL): (1) outside and (2) within Flood Peak Periods (FPP).

For comparison, I also present the interaction between (RW) and (NL) tags outside (1b) and within (2b) flood peak periods. We observe here that the varying extent of flood peak periods (FPP) has no effect on the tags' relations outside these time intervals, and that the relationship between the benchmark and the candidate lexemes are significantly stronger ($r = 0.45$, p[0.05], (1b)) than the relationships between F and RW ($r = 0.09$, p[0.05], (1a)). Figs 2a and 2b show correlations within FPP and clearly illustrate that in these instances event magnitude does matter. For the pairs of candidate- and event-describing tags, the correlation increases with magnitude, and conversely, we observe a simultaneous decrease for the tag pairs (NL) and (RW). In terms of temporal proximity to the *peaks* (i.e. 'events'), we observe that combined tags work as a better predictor one day before the flood peak events, after which the correlation drops to its minimum on the exact day of the maximum upload of UGC tagged with direct event descriptors containing 'flood' lexeme.

To investigate how each of the (RW) components performs in conditions of varying magnitude, I examined FPP only and compare relations between individual tags-candidates (R), (W) and event-designating tag (F) (Fig 4.5 (1a and 1b)) and combined tags used to describe generic environmental benchmark tags (NL) (Fig 4.5 (2a and 2b)).

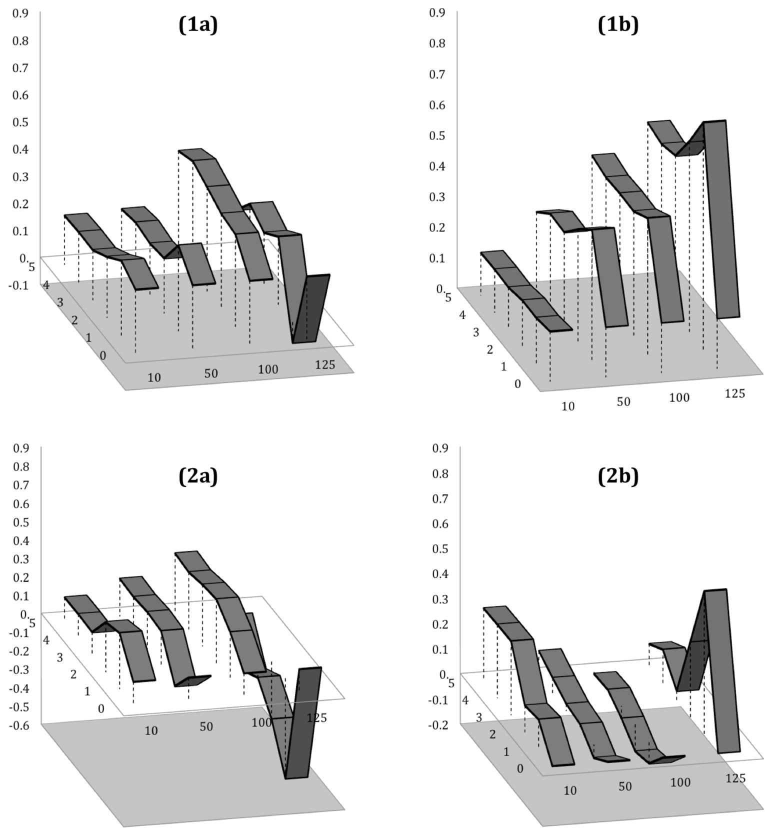

Figure 4.5: Correlation values of the cross-dependencies between deconstructed (RW) tag material: (a) (R) vs. (F) (1a) and (NL) (2a); and (b) (W) vs. (F) (1b) and (NL) (2b) within Flood Peak Periods (FPP) exclusively.

These results illustrate some slight - though significant - differences between hydrologically-related tags 'river' and 'water'. For instance, the statistical performance of the tag (R) differs very little when related to flood-signaling and natural topics alike, while the strength of the relationship drops slightly in the case of (NL) postings (from r = 0.41 to r = 0.30, p[0.05]). The pattern of this relationship also demonstrates an interesting and quite dramatic 'drop-effect', when event magnitude increases to its highest band (more than 125 uploads per day), with the correlation coefficient being at its lowest on the day preceding the actual outbreak in both cases ('flood peak'). We may also conclude that the main contribution to the correlative power between combined hydrologically-themed and flood-related tags - when the magnitude is accounted for - was predominantly due to the tag 'water'. For example, Fig 4.5 (1b and 2b) illustrates how correlation increases with event magnitude, peaking at its highest one day before the local maxima, and how the opposite to this trend is reflected in the case of longitudinal relationships with the combined (NL) tags. It is also possible to observe the highest correlation peak for the pair of tags (W)-(F) at the highest event magnitude level, with more than 125 daily postings.

Although the connections between generic and risk-signaling environmental tags have been successfully established, there is still a need to identify whether there is potential for words-candidates to be able to detect flood peaks (a.k.a. 'events') before their outbreak. For this purpose, the time window was reduced our time window to the five day pre-*event* interval and the tag relationships compared across all the best performing predictor candidates identified during the previous stages of the algorithm workflow and across event magnitudes for which they demonstrated the highest correlation capacity (more than 100 and 125 postings per day).

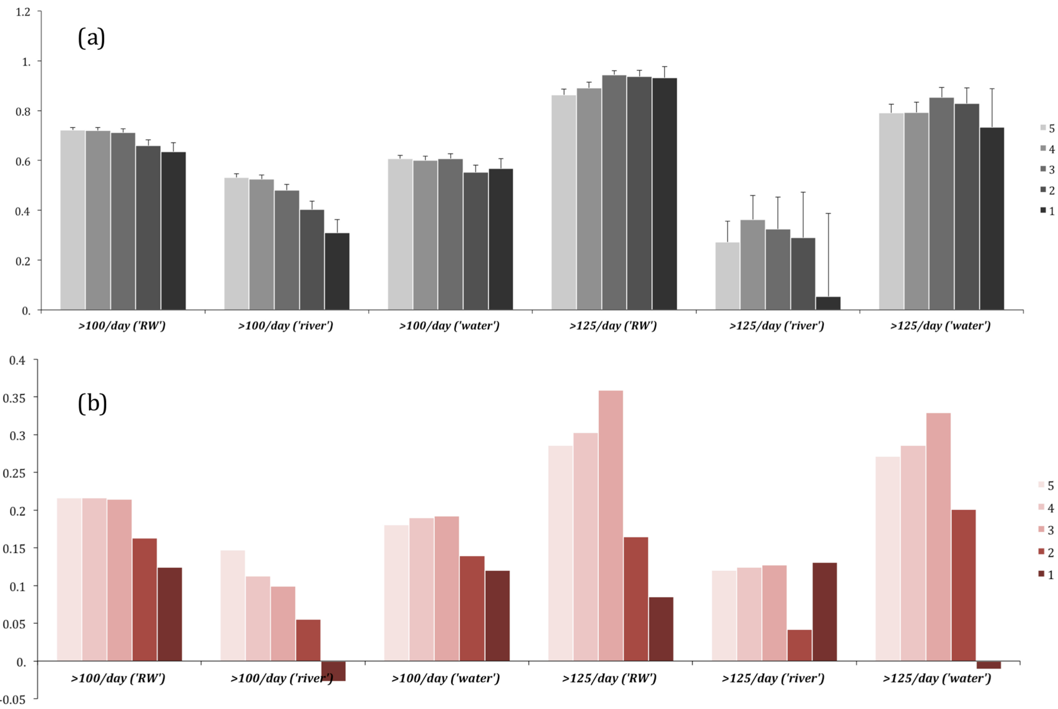

Figure 4.6: (a) PPMCC values of the cross-dependencies between the best performing predictor candidates ((R), (W) and (RW) tags) and hazard-signalling tags ('flood', 'flooding', 'floodplain') within 5-day time intervals before the event local maxima (more than 100 and 125 open source crowd-generated uploads to the *Yahoo! Flickr* platform); (b) PPMCC difference for the best predictor candidates when accounting for the pre-event interval only, as compared to the entire FPP.

Fig 4.6(a) illustrates that narrowing the analytics temporal window throws the advance prediction timing of the best candidates (RW) and (W) tags back from one to three days. Fig 4.6(b) illustrates how correlation differs between the best selected candidates during each of the five days before the event outbreak and throughout the entire FPPs, and shows a dramatic increase in the correlation power of the tag 'water' across both top comparable magnitudes (Fig 4.6(b)), thus producing a substantial peak in the aggregated (RW) tag sequence three days before the event (as an average estimate, calculated for the entire study period 2004 - 2014).

4.6 Conclusion

This data experiment, to the best of my knowledge, constitutes the only study that attempts to verify how indirect event descriptors (i.e., words that do not name the event explicitly, but are assumed to be related to it) can detect flood events when they are subjected to so called *semantic drift*, during which words change their primary meanings.

From the results I conclude that for the case of flooding taken as an event, alternative words (such as 'river' and 'water'), derived from 'social sensors', have the potential to be used as *hazard detectors*, since they exhibit meaning fluctuations that can be seen as a particular case of semantic change, i.e., *transient semantic drift*. This very simple data experiment illustrated that event-describing words can form categorical clusters, which, if selected carefully, can consist of the groups of words with polar opposite meanings, one of which can be event-describing, and words with more neutral meanings, which fluctuate between the two mentioned types. Such *neutral* words may represent the data potential for event analytics, which hasn't be exploited so far.

This conclusion was arrived at after observing changing correlation patterns with the word 'flood' around the hazard peaks and with more positive words 'nature' and 'landscape' outside those flood peak periods. These findings have a high significance and represent a substantial advance in the field of the application of social media analytics in the natural sciences, specifically as they belong to the category of *unpredictable* events [Fujiyama et al., 2016]. Further investigations in this direction are recommended in order to assist design and implementation of *socially adapted* flood warning systems, i.e., ones that integrate human behavioural data alongside hydrological and pluviometry information, however, other similar applications may also be identified and tested.

Regarding consistency with the findings presented by [Hamilton et al., 2016a], about higher semantic volatility of more negative words, these results do not provide any definite conclusions as in order to be able to achieve this level of expertise more events will need to be analysed, including social ones and the ones with much more positive connotations. What can be suggested, however, as an extension of the abovementioned experiment is more granular analysis that also accounts for the meaning-neutral words. And, of course, it would be beneficial to also include the notion of *relativity*, as while we know that there are more negative words in the dictionary and positive words tend to occur more often in everyday speech, what implications these facts have for event analytics still remains undefined.

Chapter 5

Event differentiation with spatial semantic drift

5.1 Synthesis

Although signal detection is an important property for critical data selection for complex event analytics, for the socio-natural sciences, and specifically for hazard analytics, it is important how much of that semantic signal maps *geographically* to the most precise location possible, since traditionally warning decisions are made on the basis of such spatially grounded information. This implies that we are primarily interested in words that *a priori* have much higher volume on social media as compared with direct event descriptors, and this augmented frequency is also supported by the test for *spatial statistical significance.* However, given the diversity of events within the flooding phenomena, the opportunity also appears to verify whether those words can also signify different *sub-types* of floods.

In regards to the above, I hypothesised that the processes, underlying people's activity on social media during flood events, are related to reactions and coping mechanisms with the *certain types* of floods, typical to that particular area (i.e., modifiable areal unit problem (MAUP)). These activities can accumulate over time and emerge in the data form of the statistically significant *hotspots.* The value of the additional data points, provided by lexemic drift lies in the fact that not only they increase the 'sensing coverage', but also can be instrumental in differentiating flooding types.

5.2 Background

Social media data have contributes to both *mapping* and *forecasting* of natural disasters, and is has been concluded that the combination of multiple non-authoritative data sources can help to fill in the gaps of spatio-temporal coverage of authoritative data, by using, for example, tweets [alongside gauges] as a weighting factor for creating inundation maps [Huang et al., 2018] or as part of a wavelet transform function for signal detection [Weng et al., 2011]. Recent studies have also reported positive findings, demonstrating how tweets can be used in areas lacking rain gauges or where sensors work with interruptions [Restrepo-Estrada et al., 2018]. Moreover, given the complex nature of flood events, which can be caused, for example, by surface water, excess rainfalls or groundwater sources, there is also a need to understand: **(a)** based on contiguity rules (i.e., when spatial units share a common border of non-zero length), how spatial zoning due to different configurations of hydrometric networks affect the usefulness of topical social media postings; **(b)** based on the fact that each hydrometric network is an historically evolved infrastructure (see below), which type of flooding people as 'sensors' are most likely to 'signal'.

It is given that social media activities are not evenly spread across areas [CBS, 2015]; And this may be due to various factors, including willingness or interest of communities to cover different topics, preference of one social media platform over another or turning off location sharing modules. Either way, if we ignore the last two factors as an acceptable bias, patterns of language distribution across space remains a rich topic for research, of interest for theoretical and applied linguists alike, since it can provide insights into diachronic aspects of human communication [Kavouras et al., 2005], but also provoke various interesting downstream applications, including [the topic of this thesis] how meanings behind words we use during natural disasters can be repurposed into *useful* data signals.

Emergence of word clusters in some areas and their absence in others may be due to several reasons. The primary and most straightforward one is outbreak of an event of some kind, in such cases space is usually prominently fragmented into distinct *hot- and cold spot* areas, covered by the same or similar keywords, the existence of which usually coincide with the duration of the event - unless its consequences are so marked that people continue talking about it for months. For instance, this was the case of 2013-14 United Kingdom winter floods [BBC, 2014], which not only covered substantial areas of the southern England but were also a combination of meteorological and hydrogeological conditions that led to the combination of several types of inundations and, specifically, co-occurrence of surface and infrequent type of groundwater flooding, which did not recede in Oxfordshire until the early spring in 2014 [GWF, 2014]. As a consequence, these events generated around unique 5mln Twitter posts between November 2013 and April 2014 with the hashtag '#*flood*', which on its own was not very informative as it failed to account for different types of flooding that occurred due to various water sources (e.g., tidal, surface water and groundwater ones) and generated different sets of impacts [PDF, 2011].

Another type of spatial distribution of linguistic material is less spontaneous, often requires some advanced computation, such as machine learning techniques, in order to distill recurrent patterns from the longitudinal datasets. The most common hot- and cold spots occur in urban areas (they are also known under the terms 'urban pulses' or 'metropolitan rhythms' [Miranda et al., 2016]), specifically around popular tourist destinations as they reflect technologically mediated various levels of human activities that occur at varying hourly, daily and monthly resolutions (hence the term 'beats'). Applications for this data are numerous, well-researched and constantly evolving; from real-time crime analytics and prevention [Rumi et al., 2018], to mobility tracking [Chioda, 2014] and mapping happiness [Quercia, 2014].

For spatial distributions of point data, two possible visualizations come to mind: *heatmaps* and *hot spots*. Although these terms are often used interchangeably, the differences between the two are quite significant. Whilst heatmaps are a mere representation of topological proximity between point data, hot spot analysis uses statistical analysis in order to define areas of high occurrences versus areas of low occurrence and therefore supports visualizations of indicators of statistical significance, rendering them less subjective. The designation of an area being a hot (or a cold) spot is therefore expressed in terms of statistical confidence [GIS, 2014].

Different domains that use geostatistical techniques as a method have their own understandings of what constitutes a hot spot according to underlying theories in their respective research fields, which help to avoid erroneous data interpretations. For example, crime analytics uses place, street or neighborhood theories, which are locational and are often contrasted with repeat victimization theories, which can operate at various levels of space geometries [NIJ, 2005].

In computational linguistics, hotspot analysis is commonplace although much less formalized than in the other disciplines. For example, [Grieve, 2011, 2013] used geostatistical approaches to study regional dialectology, syntactical rule deviations in Standard American English, as well as to compare its regional phonetic and lexical variations. More recent studies, by the same author and his collaborators [Grieve et al., 2017] looked at cases of lexical emergence, an irreversible example of semantic change from the point of view of geographical dimensions (contrasted to the domineering distributional semantic or longitudinal approaches mentioned in the previous Chapter of this thesis) and geolexicographical variations of common words [Grieve et al., 2011]. Whilst they are quite innovative in terms of successful cross-disciplinary methodological borrowings, these analyses also show some limitations that, when addressed, could benefit linguistic data applications, including those within the scope of this thesis. First of all, these analyses are predominantly sociolinguistic (as opposed to the phenomenological tradition). They cover a great deal of what has been previously defined by anthropological linguist Pennebaker in the 90s (and some other authors) as "mood of personality", i.e., linguistic variation due to personal styles, rather than any exogenous factors [Pennebaker, 1993; Arntz et al., 2012; Hirsh and Peterson, 2009]. Secondly, all lexemic analyses conducted with help of hot spot mappings are predominantly *univariate* [WM, 2015], i.e., they don't take into account spatial interactions of several lexical variables, which could have provided additional dimensions to temporal word vectors, used to analyze corpus-driven temporal language dynamics [LEX, 2016].

5.3 Hypotheses

According to the USGS classification of floods [FT, 2019], there are two most basic types of floods ('slow' (pluvial and groundwater) floods and 'fast' (river flow and surface water) floods). My hypotheses are therefore linked to this framework:

1. Alternative lexemes are capable to differentiate those two basic types of floods, by statistically resonating to one of them;

2. The group of 'slow' floods is better reflected on long-term social media data aggregates (in our case 2004-2014 time period).

5.4 Materials and methods

In order to address this gap I used combinations of uni- and bivariate cases of hot spot analytics in order to answer the question of whether candidate lexemes for risk-signaling do quantitatively outperform the volume of significant clusters formed by direct event descriptors (Fig 5.1). For this purpose, I tested whether candidates for semantic drift correlate with event descriptors and form statistically significant clusters outside those ones created by direct event descriptors uniquely. In order to examine their drift towards risk-signaling, I then subsequently verified whether the types of relationships they form with benchmark lexemes are significantly different.

The analysis was based on the weights of contiguity rules and modelled over the boundary geographies, derived from the main types of hydrological networks in the UK.

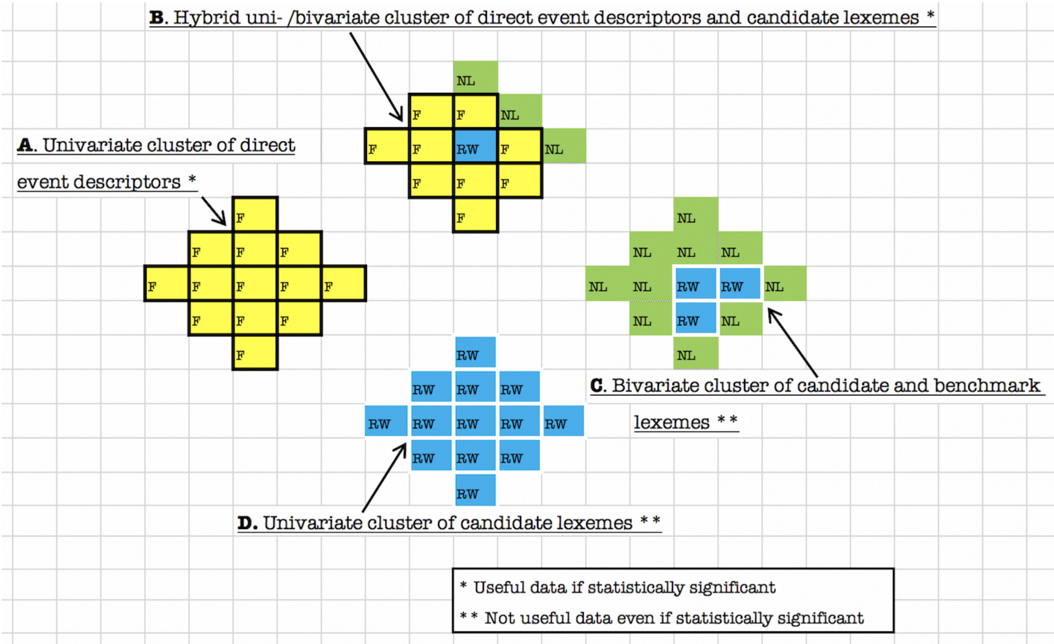

Figure 5.1: Schematic reasoning behind *useful* spatial drift of lexemic units on social media.

5.4.1 Datasets

Hydrological observation networks are composed of groups of stations or gauges and are designed to perform observations in order to maintain monitoring of a single or several interrelated objectives (e.g., an assessment of available or scarce water resources, flood forecasting, design of development plans, etc.). In most cases, networks are designed to address a set of connected objectives, such as combined flood warning and in such cases they are compositionally heterogeneous, i.e., they consist of several types of sensors (stream gauges for monitoring river levels, meteorological and agrometeorological stations for respective observations of atmospheric and soil moisture conditions, pluviometry sensors for measuring precipitation levels and intensity, etc.).

Alongside the conventional uses of hydrological information, such networks also support a set of secondary applications, including planning, drinking water quality control, biodiversity sampling, design of water utilization systems and development of local flood warning systems, just to name a few.

The design of hydrological networks and their gradual optimization is usually a continuously iterative process, often economically motivated and justified, starting from a minimum number of sensors (stations, gauges) and gradually increasing their number until the network reaches its optimal configuration. According to the Institute for Watershed Science [IWS, 2019] at Trent University in Canada [Pyrce, 2004], design of the earliest networks was primarily driven by a single, specific project, however, efforts to support wars and military operations at the beginning of the last century motivated what has been later referred to as "ad hoc installation series, without much reference to one another". Eventually, this process led to the emergence of several early observation network theories, largely experience-based, including the 'quasi-uniform areal coverage' one, which takes into account only the primary aim of installation, and what [Nemec and Askew, 1986] referred to as a 'basic pragmatic approach'.

The very first attempts to use statistical methods did not emerge until the late 1930s, when error estimates were used for choosing optimum gauge density for precipitation measures [Rainbird, 1967]. In 1934 the first network design was proposed, using elements of spatial and linear interpolation [Glushkov, 1933], thus producing continuous representation of fields of hydrological elements, which can be assessed by stations/gauges, distributed according to typical watershed areas (i.e., 'zonal-representative') in such way that they are close enough not to miss important geomorphology, but far enough so as to detect norm gradients of hydrologic elements. Each station was therefore located at the discharge outlet of the micro- (or meso-) catchments with similar characteristics, where optimum watershed area A should satisfy the relation $A_{gr} <= A_o < A_c$, where A_{gr} and A_c are the gradient and correlation criteria respectively, which should satisfy the smallest and the largest distance between the centres of fragmentation candidate catchments. Also, one of the most characteristic features of early network designs and optimization was the two-type approach to the function of stations. The first group was meant to work on a continuous basis and the secondary group was supposed to operate for relatively short time periods (up to 10 years) in order to provide benchmark calibration information for the data collected by continuously operating stations.

In the 1990s and 2000s, socio-economic strategies and information theories have been gradually introduced into the design of networks, specifically Bayesian decision theory models [Mawdsley et al., 1990; Holzkämper et al., 2012], which, amongst others, examined the economic value of data in the design of hydrometric networks for flood protection. These interdisciplinary methodological innovations are also complemented with regionalization techniques, probabilistic and deterministic models and cartographic analysis, where each method has a particular application depending on the type of problem to be solved or limitations in the collected data.

Depending on its primary purpose, the hydrometric network can be called a surface water network, precipitation network, a groundwater network or a water quality network [Mishra and Coulibaly, 2009]. Fairly recent reviews, conducted by World Bank, United Nations and WMO have noted a marked decline in hydrometric network density in developed and developing countries alike, due to reasons varying from lack of funding or appreciation of the value of long-term hydrometry, to critical infrastructure disruptions caused by wars and other disasters. As a consequence, in order to draw attention to the quality of contemporary measures of environment sustainability, WMO proposed *density of hydrological network* as one of a number of novel indicators. This is defined as the average area served by one hydrological station [Kundziewicz and Somlyony, 1997], which needs to be periodically reviewed as several researches have demonstrated the impact of network density on the accuracy of streamflow estimates. The optimization problem therefore has become two-dimensional; from one perspective the need emerged to design long term, flexibly sustainable networks from scratch, and from the other to find ways to augment a previously existing network [Pardo-Iguzquiza, 1998]. Some analyses have already been carried out regarding the latter problem, specifically in design of pluviometric networks, which successfully used variance-reduction in combination with simulated annealing [Pardo-Iguzquiza, 1998; Fattoruso et al., 2017], where estimated accuracy was maximized and the total metering cost was minimized through variance reduction and enumerative search algorithms.

In the scope of this analysis several types of the UK-wide hydrological networks have been used as a benchmark authoritative geo-designations, specifically groundwater, precipitation, surface water sensors and riverflow gauges. Their XY coordinates, along with associated stations' metadata were downloaded from the DATA.GOV.UK portal [DSP, 2019].

5.4.2 Methods

Space fragmentation

The main principle behind hotspot analytics lies in its use of vectors to identify locations of statistically significant clusters (hot spots and cold spots) in the data by aggregating points of occurrence into polygons that are in proximity to one another, based on a calculated distance. The analysis therefore groups features when similar high or low values are found in a cluster. For the purposes of the spatial statistics methods, oulined below, I used basemap polygons, derived from the *Voronoi tessellation* around each measurement station (Fig 5.2).

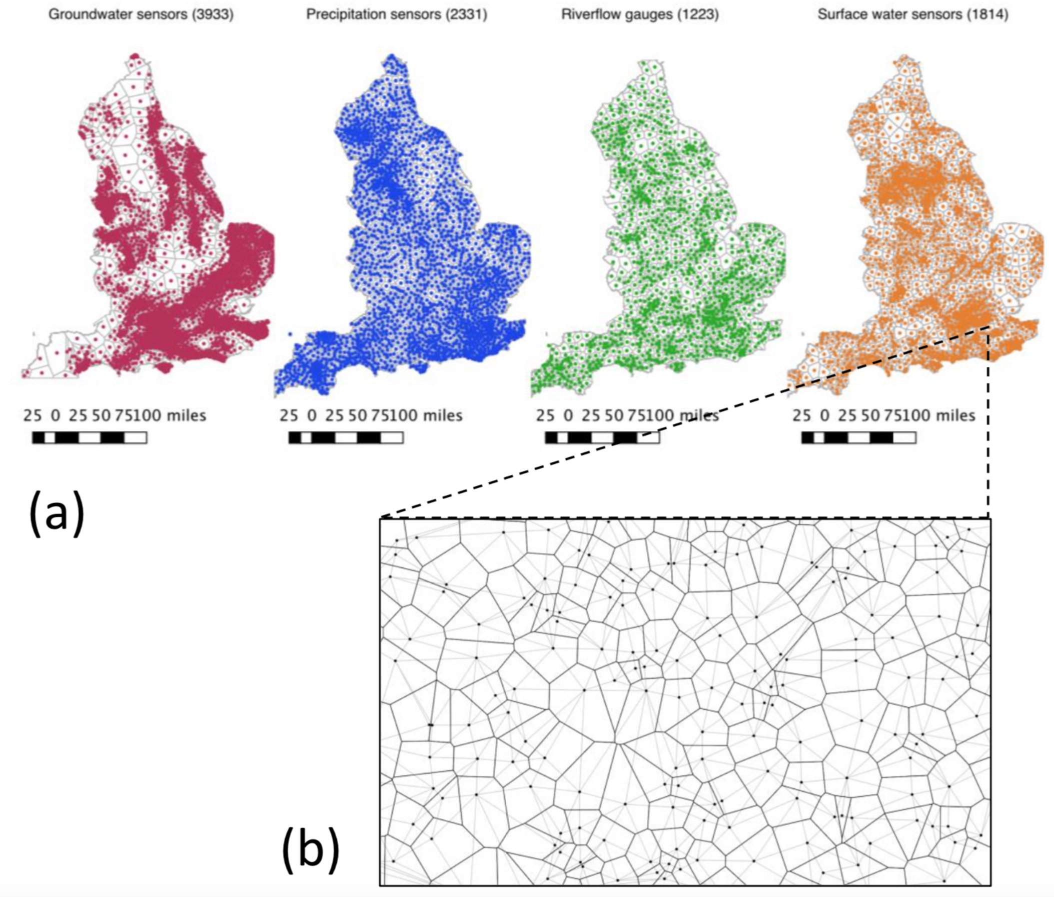

Figure 5.2: (a) Main networks for monitoring water-related risks in the UK; (b) Principle of the Voronoi space fragmentation around hydrometric monitoring points.

Global Moran's I (Spatial autocorrelation)

Before hotspot analysis is performed, it is important to test for presence of clustering in the data with some prior analysis technique involving spatial autocorrelation that will indicate if any clustering occurs within the entire dataset. For this purpose, I used global Moran's I, which belongs to the group of simple-to-use global statistical tests, amongst which are also mean center, standard deviation distance and standard deviation ellipse [Chun and Griffith, 2013]. The test for clustering is considered to be the first essential step in revealing whether data has hot spots, and there exist several approaches, such as nearest neighbour index (NNI) and test for spatial autocorrelation. All the methods start with the basic principle of hypothesis testing from classical statistics, specifically from the initial assumption that data is distributed according to the rules of *complete spatial randomness* (CSR).

Since emergence of the useful configurations of social media points is conditioned by the topological configuration of the basemap polygons (i.e., *sensors*), as well as spatial dispersion of the point clouds, the global test is used to highlight the presence of statistically significant clusters, which can be either *positive* or *negative*. Spatial autocorrelation tests, of which Moran's I is the most commonly used, require a so called 'intensity value', which is used to construct a *weights matrix* using rules of distance (between points) or spatial neighborhood directionality (polygons), and which, in case of the latter can be perpendicular, diagonal or 8-directional (Rooks, Kings and Queens, respectively [Chun and Griffith, 2013]). In the scope of this analysis I used Queens contiguity rules, however, there is also a scope for future studies to look at the distance-based sensitivity, which can be useful, for instance, to define *event boundaries* or for comparison with other spatial statistic indices.

Contiguity-based Moran's I works with point data, aggregated by polygon boundaries, where each point represents a unique attribute (in our case, each point entry represents a lexeme). The metrics of global spatial autocorrelation Moran's I therefore estimates spatial relatedness taking into account both feature locations and their values.

$$I = \frac{n}{S_o} \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_i z_i^2}, \tag{5.1}$$

where z_i is a standard deviation of feature i from its mean value, w_{ij} is the spatial weight between features i and j, n is a total number of features and S_o is the aggregate of all spatial weights.

Given sets of features that are arranged in a particular spatial configuration and are associated with a particular attribute, Moran's I measures whether the pattern of arrangement is statistically clustered (I$\sim$1), dispersed (I$\sim$(-1)) or random (I = 0) [Anselin et al., 2008]. If points that are close together have similar values, the Moran's I result is high, where the significance of the result can be tested against a theoretical normal distribution by dividing by its theoretical standard deviation [Páez and Scott, 2005].

There are several ways of drawing inferences from Moran's I and, in most cases, it is down to weights selection. A positive and significant global Moran's I suggests clustering of the like values in the dataset is present, but not necessarily hot or coldspots specifically as it could be either or both simultaneously. Indication of clustering does not provide an explanation for why clustering occurs, and therefore one can assume different processes behind the same, or very similar, patterns. In case of *true contagion*, evidence of clustering can emerge due to spatial interactions, for example, peer effects, epidemics, etc., while in the case of apparent contagion, evidence of clustering emerges purely due to spatial heterogeneity, where different spatial structures generate local similarities. Since in the scope of this experiment I am interested in the interaction between both processes, i.e., influence of the *preconfigured hydrometric layouts*, which introduces critique of the existing authoritative hazard monitoring, and the *posting intensity*, which reflects magnitudes of the public engagement, the contagion is therefore expected to occupy the *intermediate* position between its true and apparent manifestations. In order to elicit true tendencies, I also introduce the standard square mesh (1x1km) to benchmark my findings with the sensors' networks.

Anselin Local Moran's I (Cluster and outlier analysis)

While global spatial autocorrelation provides only one statistic to summarize the entire study area, the detection of clusters requires local statistics. Since Moran's I test is represented by the sum of individual crossproducts, it has been successfully repurposed for local indicators of spatial autocorrelation (LISA) by calculating individual Moran's Is for each spatial unit, associated with significance estimators. Unlike global Moran's I test, the local one (also known as cluster and outlier analysis), defines clusters of attributes with high or low values, as well as spatial outliers. The most used implementation of the algorithm is in the *Mapping Clusters Toolset* of ArcGIS Advanced (version 10.6 used here). The algorithm's outputs (local Moran's I, z-score, pseudo p-value) are available for each feature, irrespective of whether results are significant or not. The local Moran's I statistic of spatial association are therefore given as:

$$I_i = \frac{z_i}{m_2} \sum_j w_{ij} z_j, where \tag{5.2}$$

$$m_2 = \frac{\sum_i z_i^2}{n}. \tag{5.3}$$

Just as with the global scenario, a positive value for I suggests that a feature has neighboring features with similarly high or low attribute values, with which they together form a cluster. If I is negative, then a spatial attribute value is surrounded by dissimilar values and becomes an outlier if confirmed by the measure of statistical significance p.

Whether the spatial attribute is part of the cluster or forms an outlier, relationships are also determined by the statistical significance test. Some software packages have a single 95 per cent CI (the open source software package GeoDa [GEO, 2019] used here has three: 0.001; 0.05 and 0.1). The output field distinguishes clusters of high values ('High-High') or *hotspots*, low values ('Low-Low') or *coldspots*, high value outliers, surrounded by low values ('High-Low') and low value outliers, surrounded by high values ('Low-High').

Bivariate Moran's I (Spatial correlation)

As the authors of this algorithm state themselves [GEO, 2019], "the concept of bivariate spatial correlation is complex and often misinterpreted". This is mainly due to the fact that historically, the spatial aspect of the correlation was often omitted (or altogether ignored), thus leading to the assumption that the main mechanism behind the relationship between variables is the in-place correlation, whereas its implementation (GeoDa) was designed to reflect the correlation between one variable x_i and the spatial lag $\sum_j w_{ij} y_j$ of another [Anselin et al., 2002].

Bivariate Moran's I extends the idea of the original Moran scatter plot, where a variable and its spatial lag constitute two respective axes, to a bivariate context, where axes are composed of *one* variable and the spatial lag of *another*; In other words, it helps to estimate to what extent one variable at a location is correlated with its neighboring areas for a different variable. The equation below shows that the main interest of this algorithm is to visualize the slope of a regression of w_y on x.

$$I_B = \frac{\sum_i (\sum_j w_{ij} y_j \times x_i)}{\sum_i x_i^2} \tag{5.4}$$

The most common use of bivariate spatial correlations is when the variable is measured at two points in time, with the aim of understanding to what extent an observed value is correlated with its value at neighboring locations at a different point in time, i.e., to capture the *time-space cube* dynamics of one particular variable. The authors of the algorithm suggest that it is important to keep in mind that since the focus is on the correlation between x value at i location and the y values at neighboring locations, "the correlation between x and y at location i should be ignored".

In this work I implemented a special case of bivariate spatial autocorrelation applied to distributional semantics, where I considered the case of *useful* semantic drift if word-candidates, surrounded by either direct event descriptors or benchmark lexemes form statistically significant relationships of the type 'High-High' and 'Low-Low', meaning that the tendency for respective cases of hot- or cold spots' formation of the candidates with either of the polar groups of meanings (*risk* vs. *aesthetic pleasure*) determines their semantic drift. Whether this is useful for flood analytics or not is determined by the total volume of non-overlapping cases with the direct event descriptor's hot- and coldspots clusters, which typologically differ from clusters word-candidates formed by the benchmark lexemes.

5.5 Results

5.5.1 Global indicators

The first set of results is linked to the exploratory part of the analysis, aiming to reveal whether there are global patterns of spatial correlation between pairs of lexemes selected for analysis (Fig 5.3).

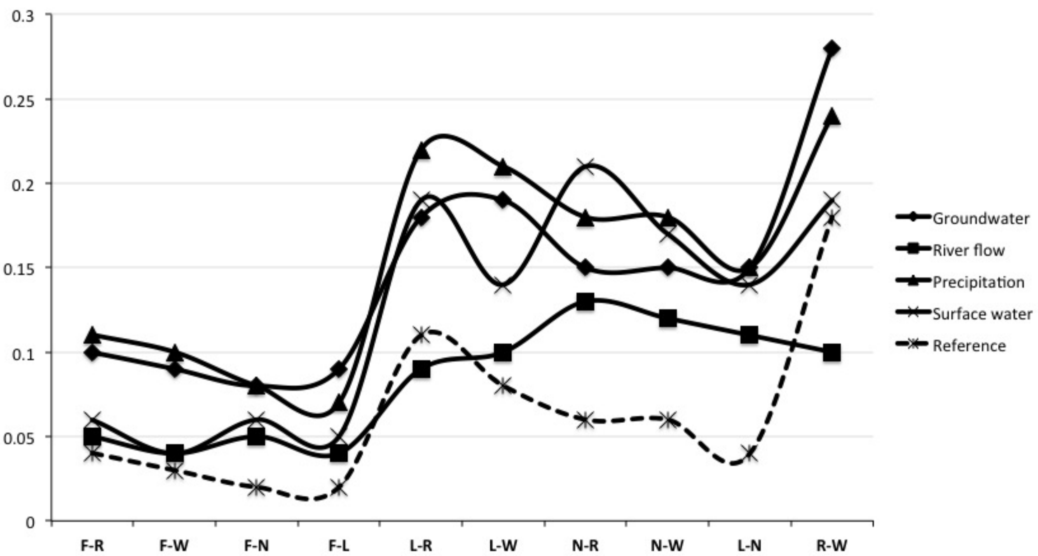

Figure 5.3: Global bivariate Moran's I testing for presence of the interactional clusters between pairs of the selected lexemes ($p < 0.05$).

The results illustrate, first of all, that globally risk-signaling lexemes (F) have the lowest tendency to form spatial clusters with either potential candidates or benchmark lexemes (Moran's Is barely exceed 0.1 at 95% CI [0.004, 0.12]), where the possibility of (N) or (L) lexemes to surround the area with the high occurrence of direct event descriptor and form statistically significant clusters or outliers is the lowest as compared across all possible combinations of lexemic relationships.

Next, we can observe elevated coefficients of significant spatial co-occurrence of the lexemes (R) and (W), which confirms their lexical dynamics in the context of this study. Both benchmark lexemes (N) and (L) illustrate stronger clustering potential with the word-candidate for semantic change (R) and lower potential with the lexeme (W), despite the fact that (W) is much more frequent in the corpus than (R) and has a much higher representation weight than its counterpart. Nevertheless, this fact could be potentially explained by the more polysemous nature of (W), which can occur in various contexts, not necessarily the environmental one or during flood events [Strang, 2004]. Similarly, the lexeme (L) forms stronger relationships with both (R) and (W) than does (N), which also hypothetically may be due to the fact that (N) may have a much broader environmental meaning and be much less connected semantically to the lexemes-candidates, whereas (L) can be directly connected to them via relations of hypernymy (where landscape constitutes a broader category of meaning into which fall (R) and (W) as more narrower, specific meanings).

In terms of the *basemap polygons* for our global Moran's I correlation tests, which are derived from the Voronoi tessellations for each group of hydrometric networks (i.e., groundwater, river flow, precipitation and reference ones [OS, 2019]), the 1km grid reference mesh prompts the weakest spatial aggregates, whilst precipitation catchments and the groundwater ones are the strongest aggregates. Interestingly, results for the river flow and surface water hydrometric networks occupy an intermediate position between precipitation-groundwater (upper threshold) and reference (lower threshold) meshes. Moran's I indices peak for precipitation and groundwater networks on the (RW) congregations (with second and third peaks, respectively, on (LR) and (LW) congregations), and for surface water and river flow on (NR) and (NW), respectively. These preliminary findings may be indicative for the potential of different word clusters to differentiate various flooding phenomena, which could be a useful source of information about the extent of flooding impacts for decision-makers and the insurance industry. Indeed, according to the USGS broad classification of floods as mentioned in the Hypotheses subsection of this Chapter, there are two most basic kinds of floods: *river floods*, which are more widespread and *flash floods*. River floods, according to this typology are more common and hence more predictable, since they occur around large rivers in areas with a wetter climate and can be linked to a storm of some kind. In contrast, flash floods occur due to excessive rainfall water, usually accumulated on top of continuous impermeable surfaces, as in urban areas, where infiltration rates are slowed down. So, looking again at the results, we can clearly spot this binary division, where one group is represented by 'precipitation-groundwater' clustering trends and is most likely to correspond to the *flash floods* type, according to the USGS classification and another group is represented by the 'river flow-surface water' clustering pattern and can be seen as corresponding to the *river floods* type. Also, if we segment the space according to the current distribution of sensors measuring precipitation and groundwater levels, then we can see that they have twice as strong Moran's I indices for lexemic clustering around direct event descriptors (F) than if we configured space according to the 'river flow-surface water'. This can be indicative of the data having stronger relevance to the *flash* type of flooding. Not only is clustering potential higher in this case, but also candidate lexemes (RW) form stronger spatial aggregations with direct event descriptors (F) than benchmark lexemes (NL). In case of the group 'river flow-surface water', it is less clear whether lexemes-candidates (RW) tend to form clusters with direct event descriptors (F) or with background words (NL).

If we add to our results visualisation the behaviour of the uni-variate clusters (FF, RR, WW, etc.), we can see that they significantly outperform the *hybrid* ones (Fig 5.4). However, although risk-signalling clusters (FF) seem to have higher tendency for autocorrelation, I needed to verify how many of the hybrid clusters (FR, FW) do not overlap with those areas, thus potentially providing some additional information.

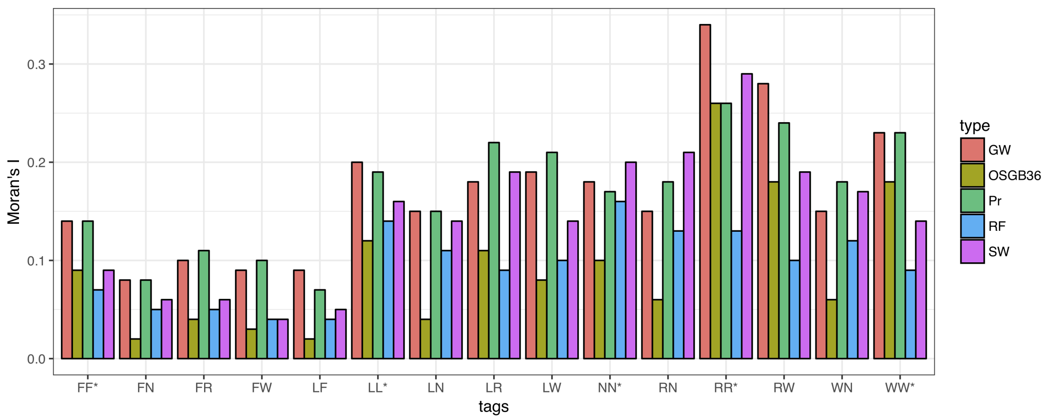

Figure 5.4: Interpretation of the Fig 5.3 when compared to the univariate (i.e., RR, WW, FF, etc.) cluster formations.

Also, I wanted to see how alternative lexemes behave in terms of the cluster formation individually and in combination (Fig 5.5). These results illustrate that when we sum up the candidates for semantic drift, they are more likely to produce stronger or more numerous clusters. This can be explained by the increase of the sheer volume of the data, also confirming the importance of continuous search of the *useful* data signals on social media.

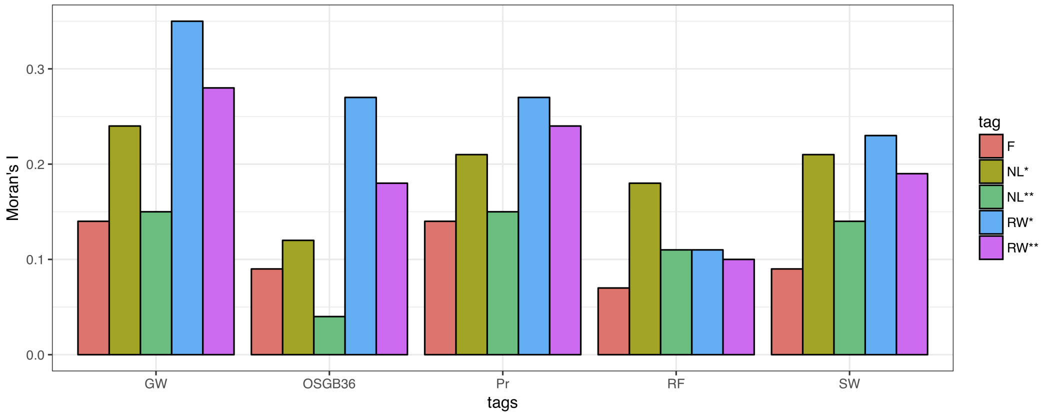

Figure 5.5: Comparison of the cluster formations when modelling neutral lexemes as independent (**) or combined (*) variables.

5.5.2 Uni- and bi-variate LISA

Once the global statistical spatial patterns were identified I then looked at local indicators of spatial autocorrelation (LISA) uni- and bivariate combinations of variables and their interactions, with the aim of answering the question of whether higher-frequency lexemes-candidates for event monitoring actually increase the amount of event-related useful [statistically significant] information for monitoring, as compared to the direct event descriptors.

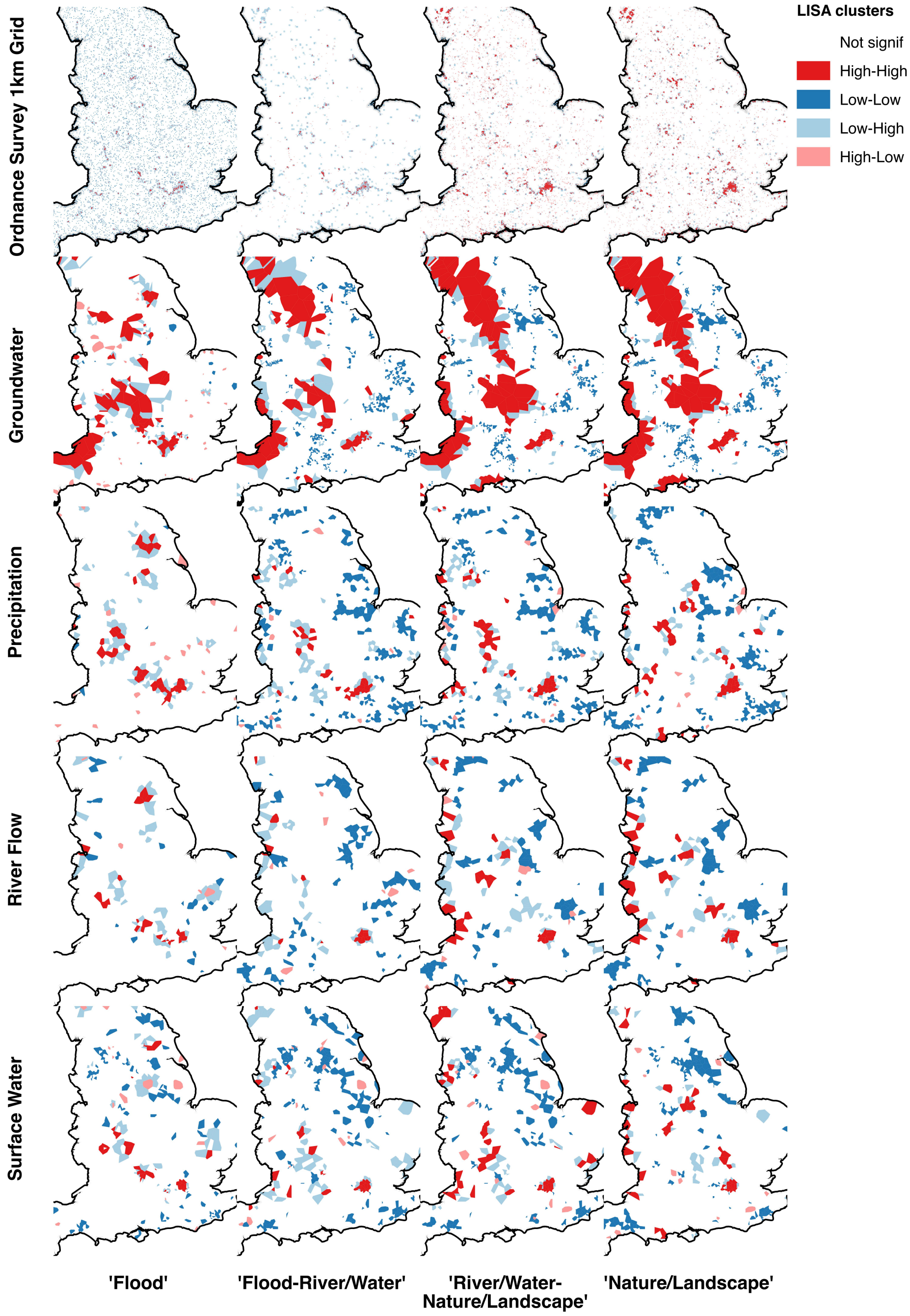

Figure 5.6: Spatial distribution of the hot- and cold spots, formed by the direct event descriptor (F) and combined benchmark lexemes (NL) (univariate cases of the local Moran's I), as well as by the spatial interaction between lexemes-candidates for semantic drift (RW) with both direct event descriptors (F) and background lexemes (NL) (bivariate cases of the local Moran's I).

Fig 5.6 illustrates the spatial distribution of hot- and cold spots, formed by the direct event descriptor (F) and combined benchmark lexemes (NL), as well as by the spatial interaction between lexemes-candidates for semantic drift (RW) with both direct event descriptors (F) and background lexemes (NL). Following the simple principle of *exclusivity*, useful data that can be produced by candidate-lexemes, will be represented by statistically significant bivariate hot- and cold spot clusters generated by direct event descriptors and the candidates, which are non-overlapping with univariate hot- and coldspot clusters, formed by direct event descriptors (hence, their additive value to already known useful variables). To avoid cases of circular spatial correlation, where candidates also form the same significant clusters with benchmark lexemes, only z-values of their bivariate interaction with the direct event descriptors (i.e., (FR), (FW) and (FRW)), which are different to z-values they form with the benchmark words ((NLR), (NLW) and (NLRW)), have been selected for each mesh cell, derived from four (plus the benchmark 1km grid) georeferenced hydrometric sensor networks.

Fig 5.7(a) illustrates the fraction of mesh cells (a.k.a. 'sensors') that are covered by statistically significant hot and cold spots of the various uni- and bivariate combinations of the risk semantics, thus suggesting the areas with the largest and smallest potential for social media posts to contribute to sensor/gauges' readings. Here we observe that bi-variate combination of risk signaling (F) and ambivalent (RW) semantics constitute the highest fraction amongst the irregular mesh cells, while representing the smallest one on the benchmark 1km OSGB grids. Predominantly across all results (apart from the FW clusters for the case of the surface (SW)) ambivalent clusters (FR, FRW, FW) quantitatively outperform the explicit risk-signaling clusters (F), thus demonstrating the additive value of semantically unstable keywords.

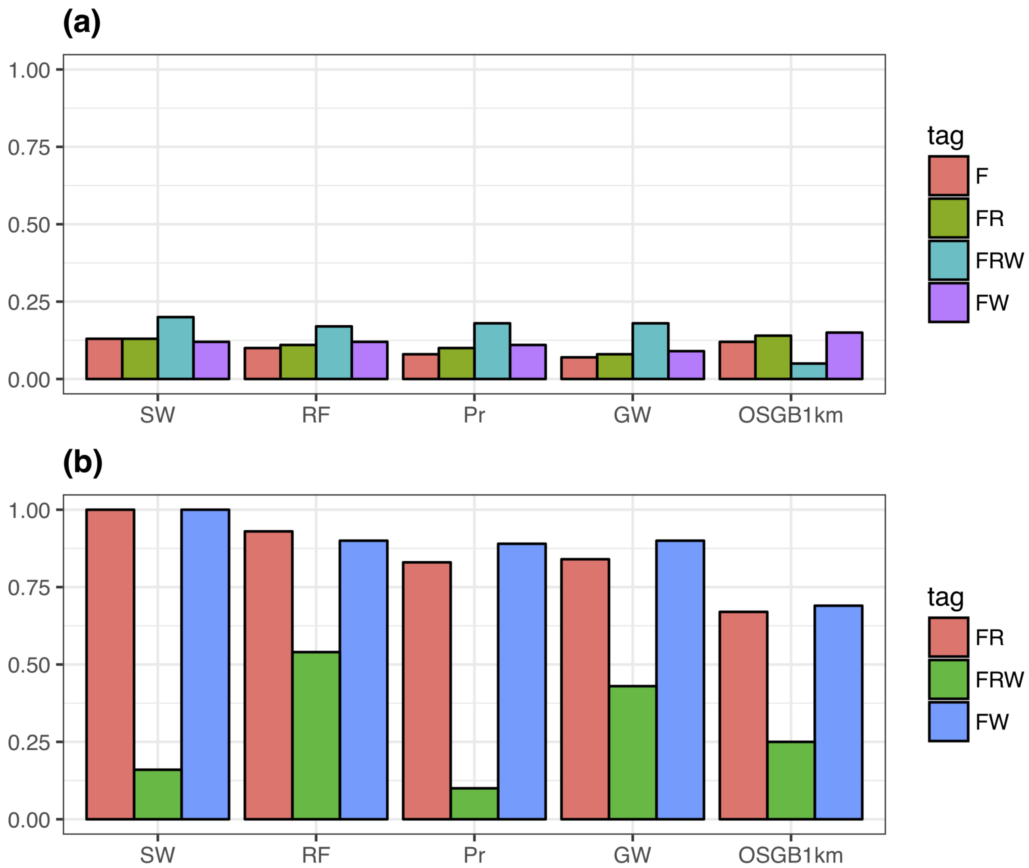

Figure 5.7: **(a)** Proportion of statistically significant units (i.e., areas around monitoring stations) to the total number of the respective units in each hydrological network; **(b)** Proportion of the statistically significant (F-RW) clusters, non-overlapping with statistically significant (F) clusters and different from the statistically significant (NL-RW) clusters.

Fig 5.7(b) illustrates the second part of the method that aimed at quantifying the fraction of statistically significant cold and hot spots (and their outliers) of the bivariate Moran's I indices that lie outside the significant (F) clusters, whilst also geographically overlapping with the bivariate Moran's I indices that (RW) form with semantically positive keywords (N) and (L), but with different z-values (thus indicating locational semantic drift).

Figure 5.7(b) illustrates the fraction of mesh cells, where statistically significant clusters of semantically neutral keywords migrated to statistically significant (and thus useful as governed by the underlying rules of spatial non-randomness) risk signaling locations. The results also illustrate the higher performance of individual keywords (FR and FW) as compared to the combined scenarios (FRW) and also demonstrate some interesting discrepancies between (R) and (W) scenarios for river flow, precipitation and groundwater networks, where (R) slightly outperforms the (W) scenario in case of river flow and where (W) outperforms (R) scenarios on precipitation and groundwater lattices.

The actual structure of the semantically drifted sensor areas is presented on Fig 5.8. Here we can observe that across all sensor networks, presence of the actual hot- and cold-spots emerges in cases of *combined modelling* of semantically neutral tags (RW). Their individual modelling only indicates the approximate positioning of the process, reflected in the domination of the positive outliers (i.e., 'High-Low' areas).

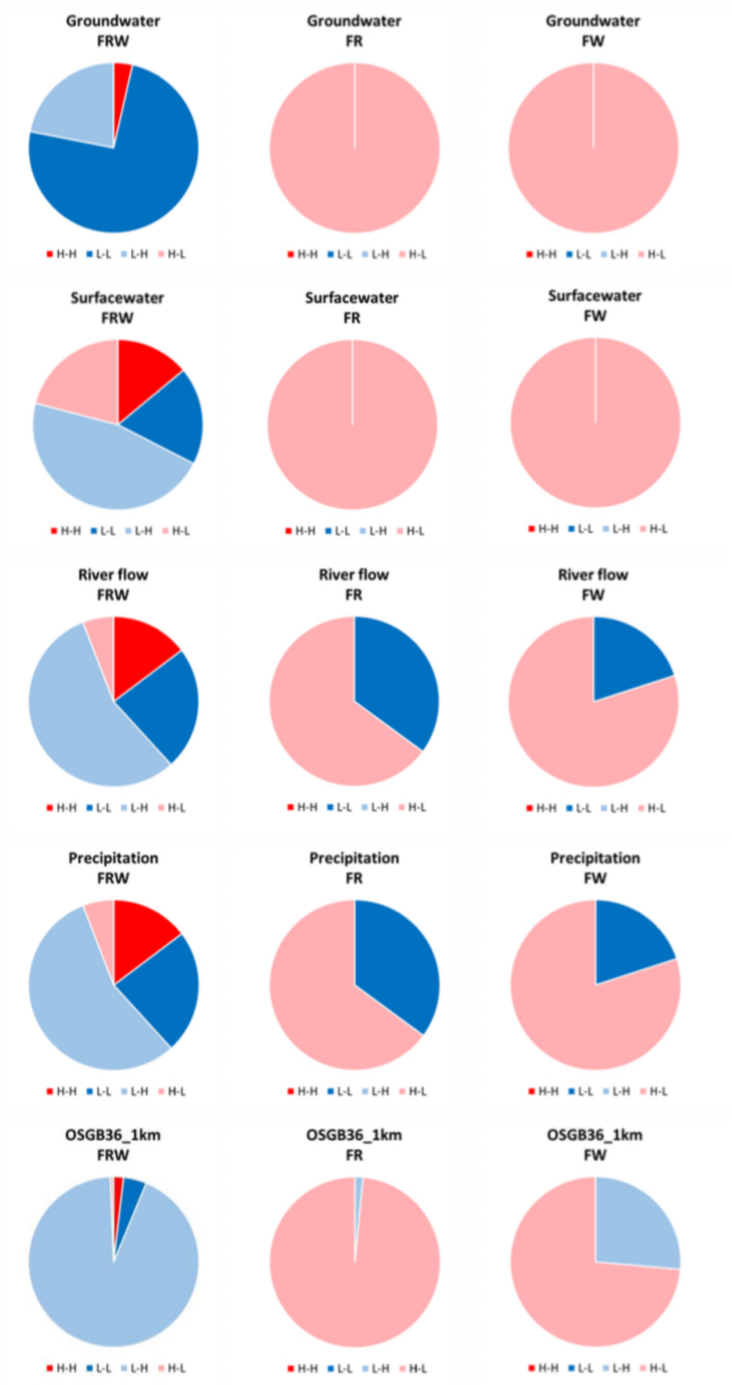

Figure 5.8: Structure of the statistically significant units (i.e., *sensor areas*, covered by geolocated human activity on social media) of the neutral lexemes or their combinations, which were observed to drift semantically from more positive environmental topics.

5.6 Conclusion

Although social media introduced a lot of geographically grounded information, applications with the combined use of the linguistic and spatial modalities are surprisingly scarce. The aim of this Chapter has been to consider some broad influences upon the process by which meaning is attributed to geographical space and regarded as being closely linked to motivation, in that the individuals invest the spatial environment with meaning according to their own experiences or needs.

In the previous Chapter I looked into how we can detect useful semantic drift on social media. Since the data we are dealing with here is firmly *geographically grounded*, the aim of this Chapter has been to demonstrate the role spatial rules play behind semantic drift and how we can extract useful data based purely on the location of polysemous UGC. Whilst pursuing the pragmatic aim of differentiating two most basic groups of flooding, I was able to confirm that semantic drift is indeed a *geographically differentiated* phenomenon and social media can help us to monitor slowly evolving or receding flood events.

Chapter 6

Exploring potential of semantic drift for event segmentation

Synthesis

As *semantic drift* is a known research category of distributional semantics for its capacity to demonstrate gradual long-term changes in meanings and sentiments of words, its empirical performance is nevertheless largely determined by the corpus composition. In Chapter 4, which used ontological relationships between words and phrases, I have already established that there also exist certain types of semantic *microchanges* on social media, emerging around natural hazard events, such as floods. My previous results confirmed that such alternative lexical material can be used to detect floods before their outbreak and to increase the volume of 'useful' georeferenced data for event monitoring.

In this final experimental Chapter I use deep learning in order to determine whether pictures associated with 'semantically drifted' social media tags reflect changes in the event severity or are a reflection of the people's reaction to the authoritative flood risk communication. The results show that alternative tags do follow the pattern of the direct event descriptors and are indeed sensitive to the evolving severity of the hazard. They also have more complex composition, ranging from more focused to less focused scenes, which can be used as statistical indicators of flood risk severity. The analysis however lacks some statistical robustness and would significantly benefit from the further testing on much larger datasets.

6.1 Background

Unusual events and changes in the natural environment can significantly impact people's day-to-day activities, therefore information on human micro-mobility has been primarily valued for its crucial role in response to disaster and evacuation strategies [Wang and Taylor, 2016]. Thus some studies reported that success of planning and executing evacuation operations to a great extent depend on exact information of where people are [Chakraborty et al., 2005; Pan et al., 2007]; Other researchers highlight that real-time designation of the risk areas could benefit from understanding the patterns of human movement [Boodram et al., 2018]. Also, successful geotargeting of appropriate shelter locations relies on hotspots of vulnerable gatherings of people [Zhao et al., 2017; Bashawria et al., 2014], whereas adaptation of early and real-time warning communication to mobile outdoor populations can be instrumental for the deployment of a new generation of *smart alert systems* [NA, 2019; Council, 2013; Gonzales et al., 2016].

Despite its obvious importance, studies looking into human mobility during natural disasters (i.e., under conditions of disruption) are quite scarce [Morrow-Jone and Morrow-Jone, 1991; Bengtsson et al., 2011]. The majority of studies seem to primarily concentrate on the fundamental characteristics of *generic* human mobility patterns [Wang and Taylor, 2016], which fall into categories of *small world behaviours* [Kleinberg, 2000], presuming existence of *cliques* and generally predictable activities, *Lévy Flights* of the exploratory chaotic movements or *Brownian navigation* associated with aggressive or predatory motives [Watts and Strogatz, 1998; Zachary, 2017; Cohen et al., 2002; Zaidi, 2012; Humphries et al., 2010; Sims et al., 2014; Karamouzas et al., 2014; Brockmann et al., 2006].

Researchers at Harvard University [Wang and Taylor, 2016] looked at 2-year human mobility data, collected from Twitter for a number of different natural disasters around the world, including hurricanes, winter storms, wildfires, rainstorms and earthquakes. This analysis was performed in order to understand whether major events can significantly perturb routine mobility patterns described by *power law* distributions. They introduced the concept of the *quantitative resilience* of human mobility, according to which it is possible to evaluate the degrees of interdependence between people's spatial movements and civil infrastructure, such that resilient activity is able to return promptly to its steady state equilibrium in response to natural hazards. They concluded that although perturbed by various hazards, the movement of people in almost all studied cases still conformed to a natural-state power law distribution, whereas event characteristics, such as *severity* and *duration*, tended to lead to much more significant disruption of urban mobility in natural hazard conditions.

While spatiotemporal data signals are useful for crowd estimation and intervention planning, research on human sensory experience during natural events is, to the best of our knowledge, nonexistent. However, as social media is gradually becoming more *visual* and less *textual* [bib, 2016; Niederer, 2018], the need also increases to adapt meaning extraction strategies from various *sensory* data modalities (e.g., video, audio, images). Specifically, for natural hazard analytics such data tranformations hold a lot of promise, since it is widely known that in situations of uncertainty people tend to generate a lot of mediated information when exploring their environment and adapting to it [Allison et al., 2006]. In this study I therefore propose to extend existing work on human *mobility resilience* and use so-called *experiential* modality, reflecting *how* people see the hazard when directly exposed to it. Building on previous findings, I focus on one natural hazard event only (flooding), however, this time accounting for the attribute of *severity* [Wang and Taylor, 2016].

Data-wise this study is also based on georeferenced (XY) information from the Yahoo! Flickr platform, where data material consists of images associated with tags and descriptive text. I made an inventory of data geographically associated with UK floodplain areas, where during the time period 2004-2014 were issued various flood risk communication messages ('Alerts', 'Warnings' and 'Severe warnings'). Selected data entries were then filtered to extract the following three categories: **(a)** direct event descriptors (i.e., tags, containing the word 'flood', **(b)** benchmark lexemes (i.e., words with which semantically unstable words highly correlate, such as 'nature' and 'landscape', and **(c)** alternative (i.e., semantically unstable) lexemes 'river' and 'water'. This framework was derived from our previous findings on the positive role of 'semantically drifted' material in flood event monitoring [Tkachenko et al., 2017a]. By analysing and comparing data across the 3-stage severity codes and before/after they have been issued, I attempt to understand how statistical indicators of the crowd attention on social media can be used for event segmentation.

6.1.1 Environmental spaces

Relationships between space and people's experiences have been well covered in [Ittelson, 1973]'s theory, where he draws a distinction between the 'space of objects' and what he termed as 'environmental spaces'. Unlike *spaces of objects*, which are usually smaller than a human body, *environments* necessitate movement within them in order to be perceived and experienced. Furthermore, unlike *object spaces*, which have little emotional content, environmental spaces also foster affective attachments, thus influencing perception of the environmental space as a whole.

Similar to this framework, behavioural geography introduced in the 1990s a distinction between *perceptual* and *cognitive* spaces. According to this, perceptual spaces refer to what can be seen or observed through the senses at one time, whilst the cognitive ones comprise larger-scale spaces, which cannot be sensed at once by our sensory system and therefore must be consecutively assembled, much like a jig-saw [Tversky, 1993]. Cognitive spaces are also considered instrumental in linking sensory images of our immediate experiences to cognitive factors of beliefs, knowledge and memory.

Since different parts of the environment are represented independently, for its successful navigation these independent representations have to be linked. Graph-like representations have long been suggested to provide a structure suitable to integrate these spatially independent, yet semantically interconnected, experiences or memories of space [Kuipers, 1978; Poucet, 1993; Schölkopf and Mallot, 1995].

In these graph-like structures, local positional information is usually attached to nodes, while edges are used to reflect the strength of the connections between them. The exact nature of information stored in nodes and edges can differ between models. Thus [Poucet, 1993], for example, suggested that nodes are place representations, while connections between distinct places are encoded as vectors in the polar coordinates of a two-dimensional coordinate system in which each point is determined by a distance from a reference point and an angle from a reference direction. Closely related to the Poucet's *network of charts* is the *theory of the network of reference frames* [Meilinger et al., 2010], which suggests that environmental spaces are represented by means of independent coordinate systems, each with its own specific orientation. Nevertheless, irrespective of structural differences, the importance of these theories lies in their efforts to structure our everyday mobility strategies according to the network theories.

6.1.2 Spatial focus

One of the ways to approach this difference empirically is to understand how people relate to the components of their surrounding environments. That is, whether they treat them as generic *objects* or distinguish them as *landmarks* [Scaplen et al., 2014; Chan et al., 2012]. According to the Burwell Laboratory of Memory and Attention [BL, 2019], there are currently many outstanding questions about the roles of 'landmarks' and 'objects' in guiding human behaviour, however, the primary difference between them lies in the fact that 'landmarks' are used for orientation purposes, while 'objects' merely contribute to the contextual background and accrue various associative properties. It has also been argued that specialisation of 'objects' as 'landmarks' should be based on the *function* of the 'object' within a specific navigational context. Where appearance is concerned, the more distinct an 'object' looks within an environment (and more informative or memorable its features are), the more likely can be associated with the 'landmark' category [Stankiewicz and Kalia, 2007]. Also, the *stability* of 'objects' in the environment can influence their role as 'landmarks'. If the former are to be counted as 'landmarks', then they need to be able to provide reliable navigational information, predominantly at the expense of a stable spatial position as it has been previously shown that objects at decision points are better remembered than those at non-decision locations [Jansen-Osmann and Fuchs, 2006; Kessels et al., 2011]. A study of virtual route-navigation by [Mallot and Gillner, 2000] demonstrated how objects in the environment attain action-related associations, and although 'landmarks' are commonly referred to as discrete objects, the geometry of their extended surface or boundaries can also provide important information for navigation.

6.1.3 Semantic density of scenes

During our movement in space, a sense of direction can help us to establish an understanding about spatial relationships between different locations and can improve the representational stability of situated real-world objects [Wang and Spelke, 2000].

For humans, orientation and directional information are controlled predominantly visual cues and hence it can be argued that for successful navigation in space one needs to operationalise already accumulated storage of visual information about previously visited locations or to create new mental images for current or future references. Performance for aligned versus misaligned (or *connected* vs. *disconnected*) orientations is therefore considered to reflect the fact that semantic relationships between objects or scenes in the real world are assigned similar connections in memory with respect to the specified reference direction.

6.2 Statistical indicators of crowd attention

Spatial attention belongs to those fundamental behaviours that are essential for everyday life. It serves the purpose of survival and involves nearly all sensory systems, though visual information appears to prevail while traversing the environment in a purposeful manner [Waller and Richardson, 2008; Murray and Wallace, 2011; Mast and Jäncke, 2007]. Also, as a research branch of behavioural geography it is primarily concerned with the question of how spatial information such as *orientation (or* *direction)* and *attention (or focus)* are coded cognitively. Specifically, whether this is done *egocentrically* (i.e., in direct relation to the observer as a primary reference point) or *allocentrically* (when the reference is a visual frame situated independently from the observer's position in space). The question of what is the difference between the strategies from the perspective of their cognitive underpinnings remains a topic of debate [Ekström et al., 2014].

In order to deploy the concept of experiential mobility, I decided to extract statistical indicators of the crowd attention from the geotagged images, taken during the sequence of flood events 2004-2014 in the UK. Such indicators belong to the category of the 'natural scene statistics' and are quite common application in the field of computer vision [Stansbury et al., 2013]. Such indicators are: (1) descriptive (qualitative); (2) quantitative and (3) relational.

6.2.1 Hypotheses

Analysis of perceptual experiences during flood events can be sensitive to both spatial and temporal design constraints. For example, we can characterise public behaviour during individual events, behaviours for a particular area across multiple events or across events of the similar types of severity. Whilst these questions would form a nice exploratory analyses for subsequent case studies, for the purpose of this thesis I decided to follow the route aiming to make sense of public behavioural response to event evolution and flood warning information [Goulter and Myska, 1987; Parker et al., 2009; Du et al., 2016].

In order to evaluate visual experiences of the exposed crowds, I used spatial designations for each of the 3-stage flood risk communications (i.e., 'Alerts', 'Warnings' and 'Severe warnings'). I selected images from the Yahoo! Flickr platform that are tagged with either of the three groups of keywords: (a) direct event descriptors (i.e., 'flood'); (b) *alternative* lexemes, exhibiting transient semantic drift around flood events ('river', 'water') [Tkachenko et al., 2017a] and *benchmark* lexemes, used to describe the general, undisturbed state of the natural landscape of floodplains (i.e., 'nature', 'landscape'.

My interest to structure the method around semantic drift in this analysis is twofold: (a) First of all, I am interested to observe the phenomenon dynamics (both qualitative and quantitative) at much finer, *sub-event*, scale; (b) Secondly, it would be useful to verify the usefulness of different groups of lexemes for various flood stages. And since the three abovementioned groups of lexemes are treated here as situational construals that are reflected in the visual and linguistic modalities of the dataset [Divjak et al., 2016; Reiter and Sripada, 2002; Garcia et al., 2012; Rohrdantz et al., 2012], my two main hypotheses are as follows:

Hypothesis 1: Scene compositions of the direct event descriptors and semantically drifted material start resembling each other as event evolves;

Hypothesis 2: Since semantic drift is a context-driven rather than controlled process, we should expect lexico-visual responses to be more sensitive to event severity when compared to the timing of flood risk communication.

6.3 Materials and methods

We already know that people tend to classify scenes according to their context and compositionality, which means that natural scenes can relate to each other as semantically close or distant in a manner similar to the principle of linguistic word embeddings. According to the latter, words or concepts tend to form situational clusters that do not necessarily coincide with their ontological (or dictionary) similarities, but do so according to their uses and co-occurrences in various domains of everyday life. With this concept in mind, I extracted computer classified scenes, alongside their classification confidence values, which will serve as indicators for the definition of 'objects' and 'landmarks', and verified how they relate to each other semantically within floodplains associated with various degrees of risk.

6.3.1 Datasets

YFCC100M

I used the Yahoo Flickr Creative Commons 100M (YFCC100M) dataset [Thomee et al., 2016] containing a list of images and videos uploaded to the Yahoo! Flickr platform between April 2004 and August 2014. All the audio-visual material provided in this database is licensed under one of the Creative Commons copyright licenses (CC:BY).

Flood stages and risk communication

Flood stages are used to describe the progress in covering the designated flood risk areas with water. The main principle behind the designation of flood risk areas is topographic gradient [Tewolde and Smithers, 2006]. Originally derived from direct geodetic surveys, now floodplains are designated with the help of more dynamic remote sensing techniques, using repeat high resolution orthophotography and photogrammetry.

Designations of topographically defined flood risk areas are used for various purposes. For example, insurance companies use them to automatically identify at-risk properties. Also, depending on the flood stage progression, flood risk areas are used by authoritative environmental bodies (such as the Environment Agency in the UK) to inform the public and organise rescue and evacuation campaigns.

In the UK, there are three types of risk communication messages, corresponding to the stages of event severity: Alerts ('*Flooding is possible. Be prepared*'), which are used from two hours to two days in advance of flooding, Warnings ('*Flooding is expected. Immediate action required*'), which are used from half an hour to one day in advance of flooding and Severe flood warnings ('*Severe flooding. Danger to life*'), which are put in place when flooding poses a significant threat to life.

Spatial designations of floodplains under Alert, Warning and Severe warning statuses and historic records of risk communication are available from the Government Data Portal [DSP, 2019]. The spatial intersection of these areas with the Yahoo! Flickr posts is illustrated in Fig 6.1.

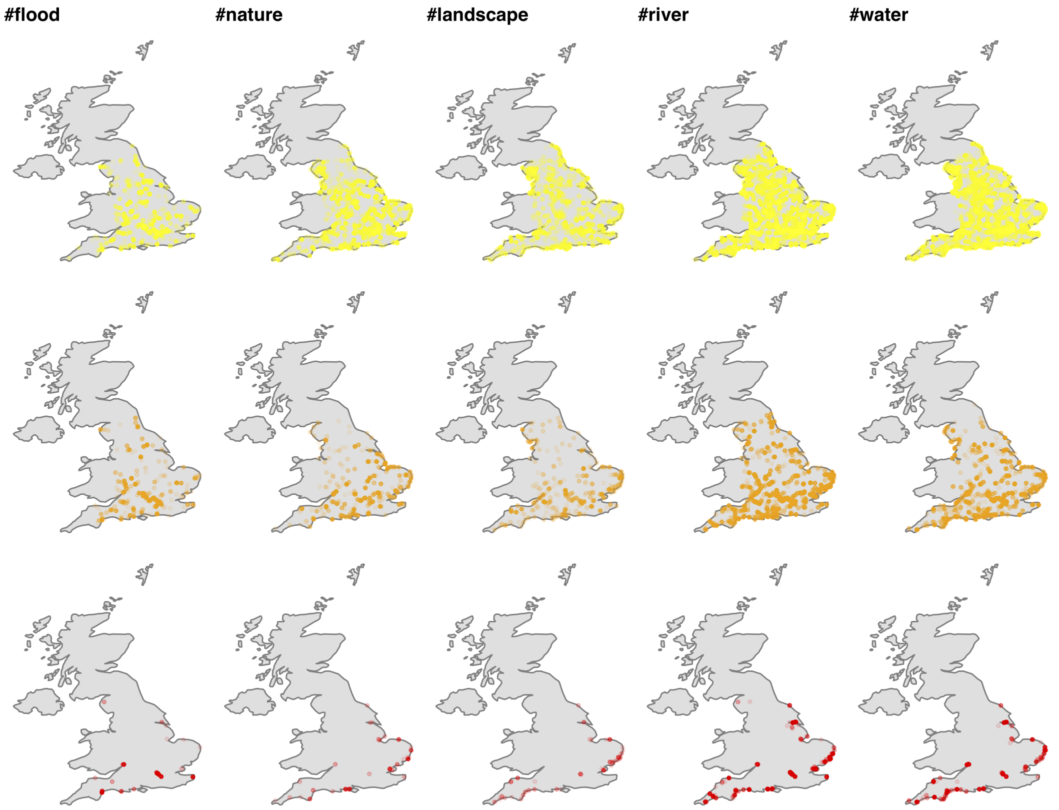

Figure 6.1: Spatial extraction of social media data. Distribution of geolocated Flickr tags uploaded to the platform during 2004-2014 within spatial designations used as communication units for *flood alert* (yellow), *warning* (orange) and *severe warning* (red) messages by the Environment Agency, UK.

For designations of the 'before' and 'after' periods around flood risk communication, I selected 100 hours (appr. four days) in each direction around the timing of the announced risk status for each designated floodplain. The temporal distribution of relevant tags around ensembles of flood events 2004-2014 is illustrated below (Fig 6.2).

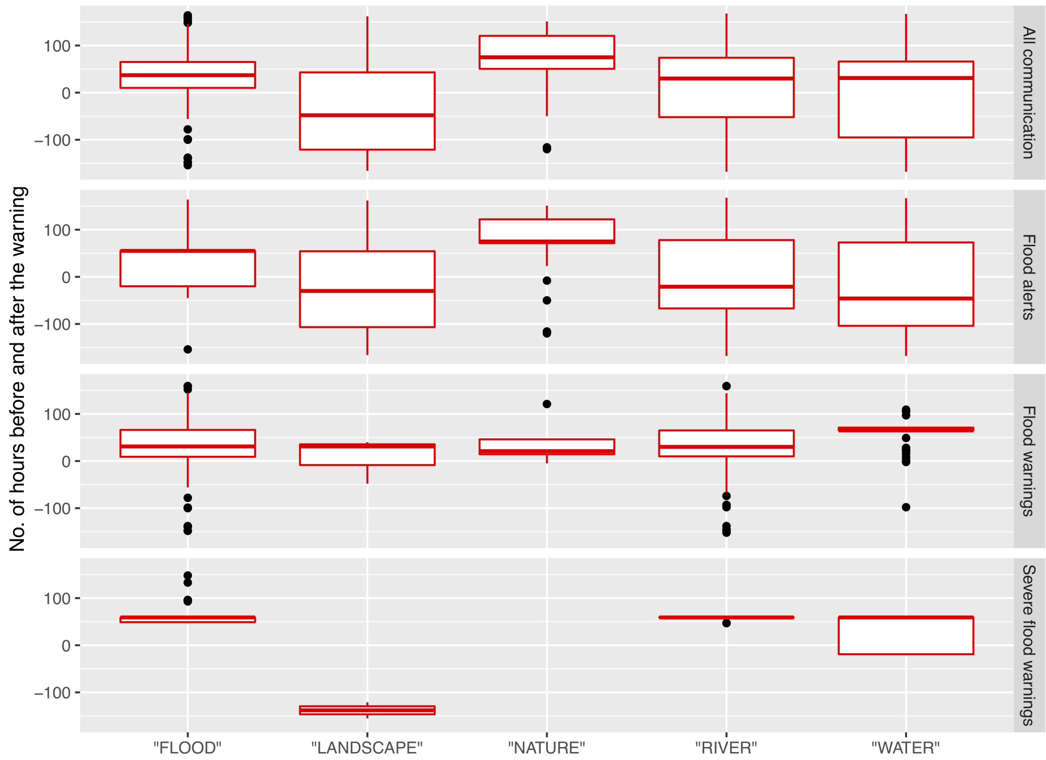

Figure 6.2: The temporal distribution (±100 hours) of tags around announced major flood events in the UK (2004-2014), using 3-stage flood risk communication system (alerts, warnings and severe warnings).

6.3.2 Methods

There are two types of information we need to extract from the social media dataset: **(a)** classification of scenes into categories of 'objects' and 'landmarks' posted before and after flood risk warning messages across all three stages of event severity (alerts, warnings and severe warnings), and **(b)** semantic relatedness of identified 'objects' and 'landmarks'.

In the case of the allocentric crowd attention there are no obvious landmarks in sight, as well as no obvious connections between places. So, whilst adapting this statement to the properties of our data, it can be argued that *landmark-equivalent* corresponds to the *well-defined scene* associated with the highest probability value by the classification algorithm [Stansbury et al., 2013]. Connections (or their absence) between places can be also expressed with the help of the statistical probability of co-occurrences of scenic categories near each other, for example, in news outlets, which comprise a substantial topical corpus on natural disasters due to their ‘newsworthiness’ [Gold, 1980]. In the case of the egocentric collective attention, the situation is the opposite, where we should expect an increased number of well-defined, typical scenes with strong semantic connections.

This framework is quite flexible as it can be adapted to multiple questions concerned with spatio-temporal, cognitive dynamics of crowds (Fig 6.3).

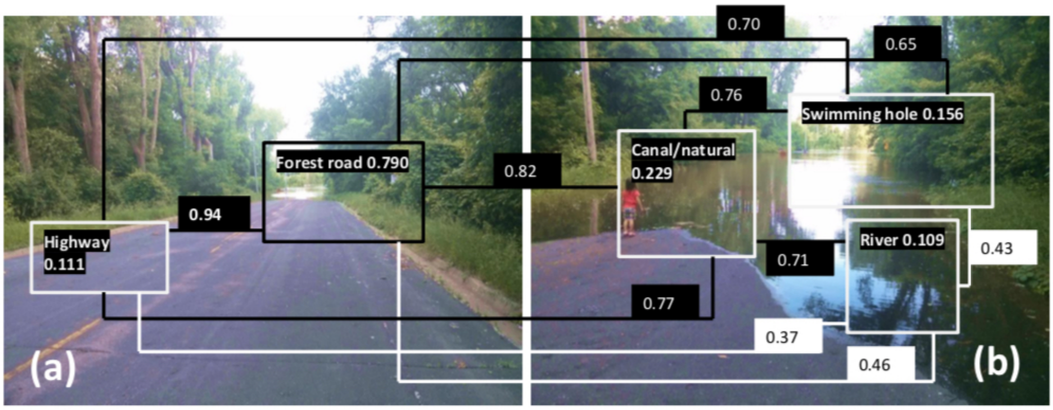

Figure 6.3: Allocentric-egocentric scene attribution. Flickr images (‘Flooded Street’ by pelennor – CC BY-NC-ND 2.0) posted during the same hour from the same area. The scheme demonstrates the gradual shift from egocentric to allocentric scene compositions.

‘Deep’ image classification into ‘objects’ and ‘landmarks’

For natural scene classification I used the pre-trained Places CNN from MIT [Herranz et al., 2018; Zhou et al., 2014], which classifies images into 365 scene categories. This dataset was designed to account for the human visual cognition system and is widely used for training classifiers to recognise high-level visual tasks, such as object detection, scene classification or event prediction. Each scene category is described with a two-tier labeling system, where simple nominal semantic categories (such as ‘road’ or ‘forest’) are associated with their functional counterparts (e.g., ‘broadleaved forest’, ‘mixed forest’, ‘city road’ or ‘desert road’). Following this classification, each image was allocated up to five scene categories and each of these values were used to make a decision whether the classified scenes corresponded to the categories of ‘objects’ or ‘landmarks’.

Jaccard distance

As this analysis is conducted across ensembles of spatial units (i.e., floodplains) and temporal segments (‘before’/‘after’ events) I used the metrics of compositional dissimilarity across extracted spatio-temporal groups of images. For this purpose we chose *Jaccard distance* [Jaccard, 1912], which reflects the degree of dissimilarity between situational scene ensembles and, in our case, aims to illustrate whether people tend to focus on the same or different areas during the various stages of flood events.

$$d_J = \frac{A \cup B - A \cap B}{A \cup B} \tag{6.1}$$

Semantic density of complete graphs

It can be argued that since environmental spaces require 'panoramic' observation to be effectively perceived [Sweeny and Whitney, 2014; Freundschuh and Egenhofer, 1997], the scenes-snapshots they are composed of also possess some kind of semantic interaction, due to crowds' attention, for example, to important aspects of flood events (e.g., dramatic scenery of flooded houses and gardens, submerged vehicles, etc.). Therefore we can use *interactional* methods for their estimation. Here, I decided to turn to fundamental graph methods, which aim to explore semantic relatedness of scene clusters posted around each type of flood event (moderate ('Alerts'), severe ('Warnings') and dangerous ('Severe warnings')), in order to visualise semantic pathways between previously identified 'objects' and 'landmarks'. We therefore can observe that a complete model of spatial navigational behavior for the area A during the time interval $(t_1 - t)$ resembles the shape of a complete, weighted graph $G\ (E, V, w)$, where $w{:}E \rightarrow eVal$ and $eVal$ represents set of potential graph weights.

It can be argued that this type of situational *semanticity* can be analysed with the help of traditional embedding methods, where the model is usually powered by the domain-specific corpora and is used to extract semantic weights between lexical items (names of the scenes in our case) based on their co-occurrences. Following this principle, we used a standard *word2vec* cosine similarity algorithm for weights compilation, where semantic similarity between two lexical concepts A and B is represented as:

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum\limits_{i=1}^{n} A_i B_i}{\sqrt{\sum\limits_{i=1}^{n} A_i^2}\sqrt{\sum\limits_{i=1}^{n} B_i^2}} \tag{6.2}$$

My algorithm was based on a pre-trained Google word2vec model [w2v, 2013] containing three million words and phrases, which has been trained on Google News data (around 100 billion words) and fitted using 300-dimensional word vectors (features).

Finally, I estimated sets of graph densities to be compared with each other using proportions between actual and potential semantic weights, where 0 means that scenes are semantically unrelated (conditions of poor or lack of orientation) and 1 illustrates topically connected clusters of the natural scenes:

$$d_G(A, \Delta t) = \frac{\sum(w : E)}{\sum eVal}. \tag{6.3}$$

6.4 Results

6.4.1 Compositional dissimilarity

First I decided to look into how scenes tagged with alternative (*neutral*) lexemes ('river', 'water') differ from the two other groups of risk signalling ('flood') and benchmark words ('nature', 'landscape') (Fig 6.4).

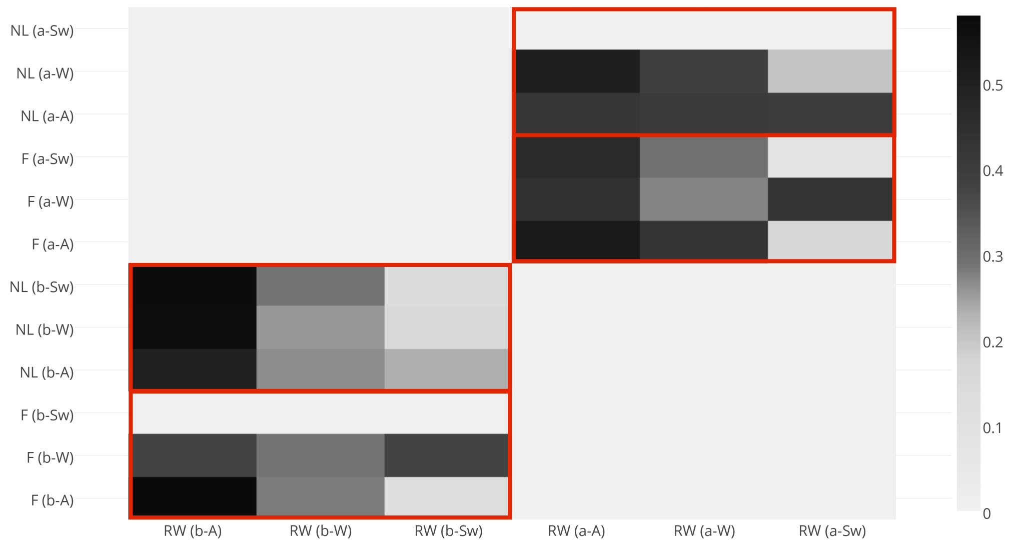

Figure 6.4: Jaccard distance between the scenes, tagged with neutral, positive and risk-signalling words posted before and after flood events 2004-2014. These results illustrate that the compositional distance between neutrally tagged photographs and the two other sets generally decreases with event severity, both before and after risk communication messages. Zero values here correspond to 'no data'.

Here, neutral lexemes, which have previously demonstrated a transient shift of meaning around flood events [Tkachenko et al., 2017a], show an increased structural dissimilarity with both sets of words. This distance gradually decreases with the increase of event severity, for both cases before and after official risk communication messages. This can be indicative of the fact that during the early stages of flood events alternative lexemes are associated with more generic natural scenes and as the hazard evolves, their similarity with the 'flood'-tagged scenes increases. However, this step required subsequent comparison of scenes tagged with event descriptors ('flood') and positive words ('nature', 'landscape') (Fig 6.5).

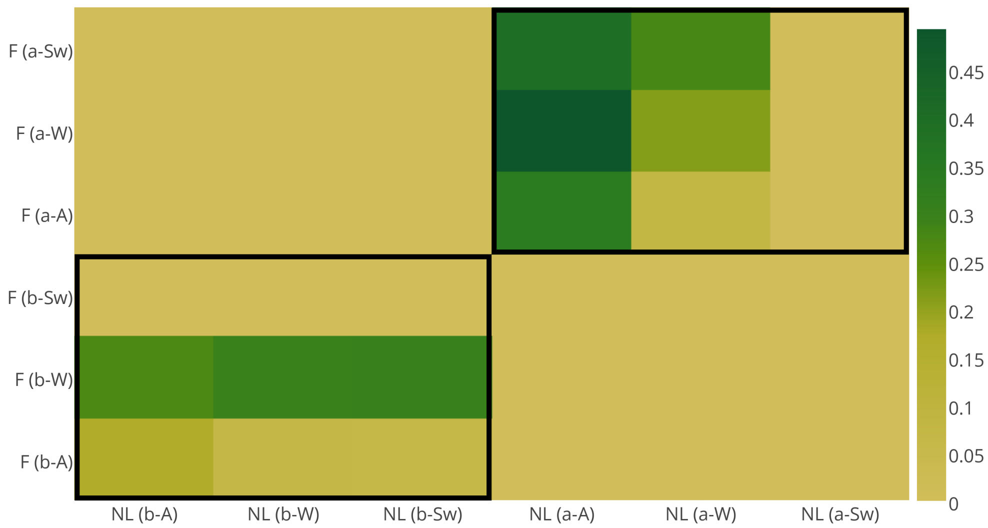

Figure 6.5: Jaccard distance between scenes, tagged with positive ('nature', 'landscape') and risk-signalling ('flood') words posted before and after flood events 2004-2014. These results illustrate that before risk communication the dissimilarity increases evenly between 'flood'-tagged and positively-tagged scenes with the increase of the event severity. After risk communication it also evenly decreases with event severity. This can be indicative of the fact that the perceived event severity affects segregation of the visual material in the same manner as authoritative risk communication, where the former segregates crowds according to the perceived danger, whilst the latter re-focuses their attention back onto familiar landscapes.

Here we can observe that the compositional distance of positively tagged scenes posted before flood risk communication, varies very little with event severity and this pattern replicates for the 'flood'-tagged scenes after authoritative warnings. This suggests that 'flood'-tagged scenes hold the potential to discriminate the severity of evolving flood events before risk communication, whilst positively tagged material have the potential to indicate post-event recovery when analysed alongside each other. However, definitive conclusions are difficult to draw because of the lack of 'flood'-tagged material posted before severe warnings and positively-tagged scenes after.

Finally, I looked at the compositional distance that the same three sets of lexemes tend to exhibit between themselves before and after authoritative risk communication (Fig 6.6). The results show the biggest structural distance in case of the *alternative* lexemes ('river', 'water'), and the smallest for the case of positively-tagged scenes 'nature' and 'landscape', with risk-signalling material occupying a somewhat intermediate position between both groups.

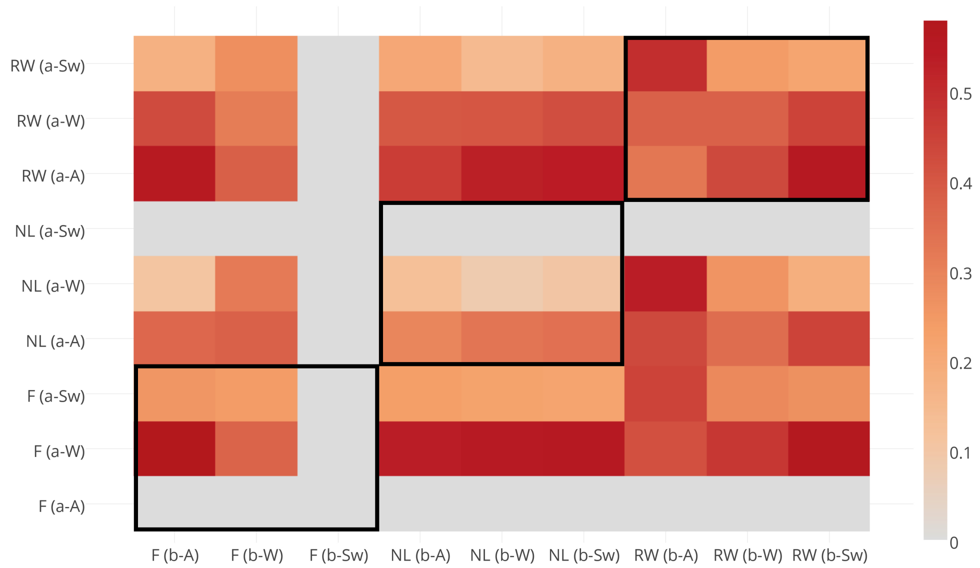

Figure 6.6: Compositional Jaccard distance between the sets of images posted *before* (horizontal axis) and *after* (vertical axis) ensembles of flood events 2004-2014. The results illustrate that scenes tagged with the *alternative* lexemes-candidates for situational semantic shift demonstrate the highest compositional distance before and after flood risk communication, which is also independent of the event severity.

Looking at these sets of results it is therefore possible to conclude that event-specific semantic drift of the neutral words ('river', 'water') discovered in our previous work [Tkachenko et al., 2017a] is also supported by the compositional dissimilarity of the images with which they are associated. Despite their temporal correlation with both sets of lexemes (positive and risk-signalling), the structural dissimilarity of their associated scenes across both sets – which decreases with event severity – may be indicative of the discriminatory potential for the severity of evolving hazards before authoritative risk communication takes place, as well as of varying (according to event severity) coping mechanisms of crowds after formal announcements of risk states.

6.4.2 Quantitative and relational statistics

Spatial focus

Fig 6.7 illustrates results of the deep learning image classification with help of the pre-trained Places CNN (See Methods section above). Here we observe very little variation between the strengths of scene classification across all three groups of images associated with positive ('nature', 'landscape'), negative ('flood') or neutral ('river', 'water') semantic tags. Nevertheless, it was possible to extract sub-populations with the classification confidence above and below 0.5 in order to perform some further relational analysis of the semantic density between, what we can define as 'more focused' (p>0.5) and 'less focused' (p<0.5) scenes.

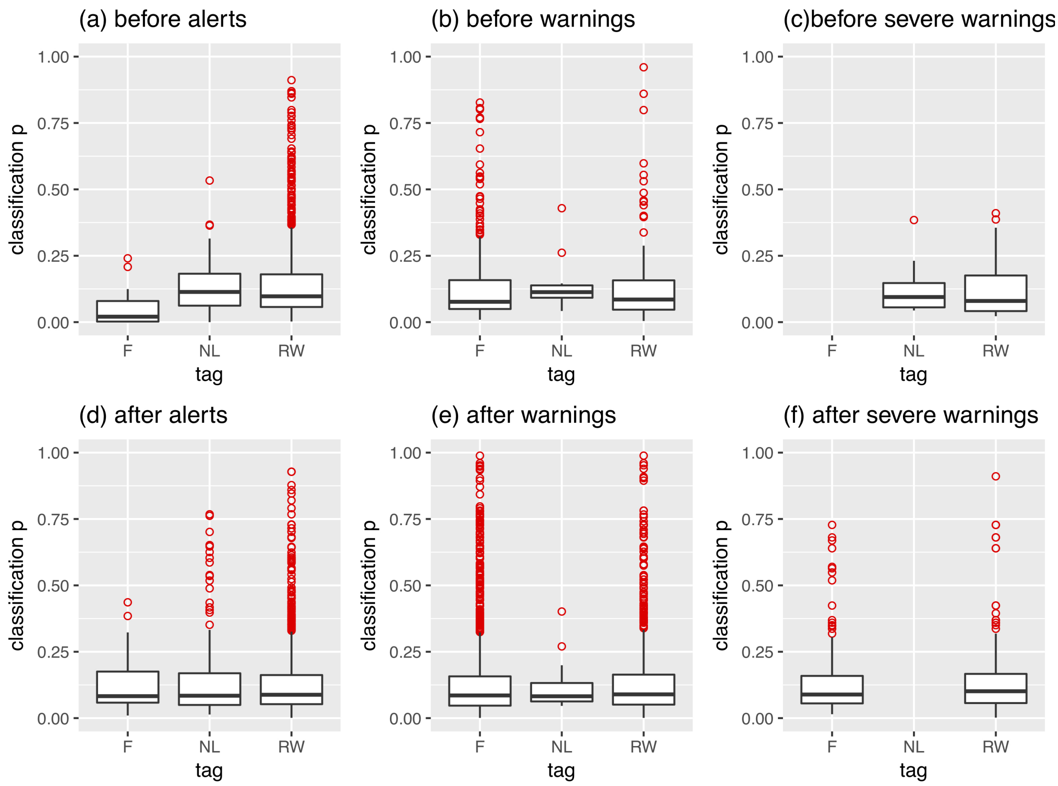

Figure 6.7: Distribution of the confidence values of the natural scene classifications using MIT Places CNN.

Semantic density of scenes

Semantic density is a network-based statistical indicator, based on the lexical properties of the scene classification step. Using word embeddings and lexical network analysis (Chapters 3.4.3 and 3.4.4) I explored relational density between the lexical categories, which are used to describe scenes with the different degrees of focus. Fig 6.8 illustrates the following findings:

1. The semantic density of all 'flood'-tagged scenes gradually decreases with the increase in event severity. Following previous sets of findings, this phenomenon is coupled with the simultaneous increase in spatial focus. In the case of semantically unstable material the trend is exactly the opposite: increased semantic density is accompanied with a decrease in focus;

2. Amongst groups of outliers and as compared to the entire datasets, the most dramatic examples of semantic density are for scenes tagged with semantically unstable words and this density also increases with event severity. It is therefore possible to conclude that after emergence of 'flood'-tagged scenes, the rest of the *alternative* lexemes start losing their significance as *risk-signallers* and prepare to *mutate* back to more positive connotations (i.e., 'nature' and 'landscape').

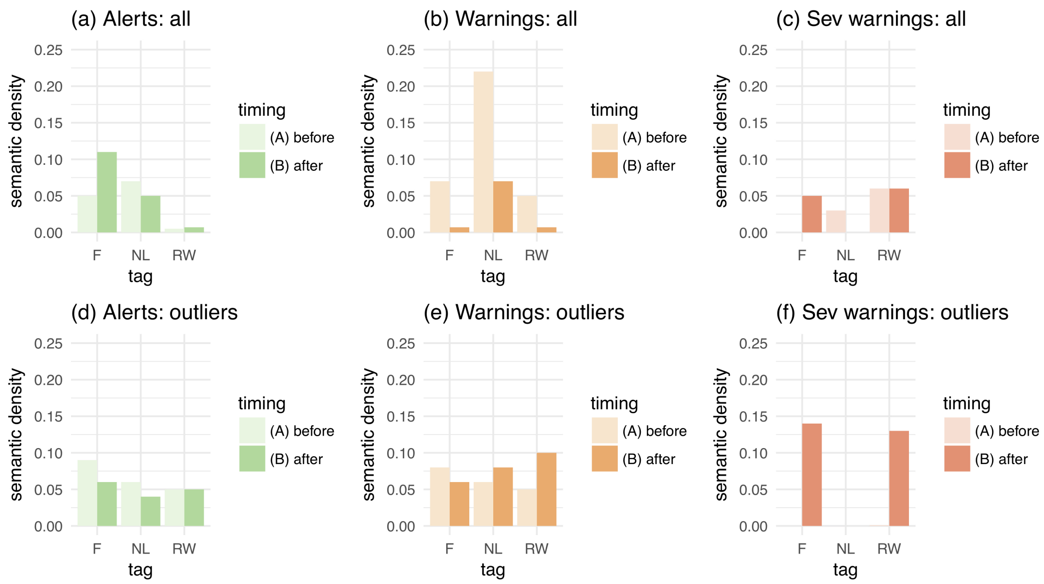

Figure 6.8: Semantic density of navigational frames captured by images on the Yahoo! Flickr platform posted before and after official flood risk communication.

6.4.3 Cross-interval comparison of focus distributions

To finalise the results obtained during data exploration I performed Kruskal-Wallis H Test on the spatio-temporal data segments. Prior to this I run the normality test (Fig 6.9), rejection of which suggested use of the test for population distributions without assuming them to follow the normal distribution.

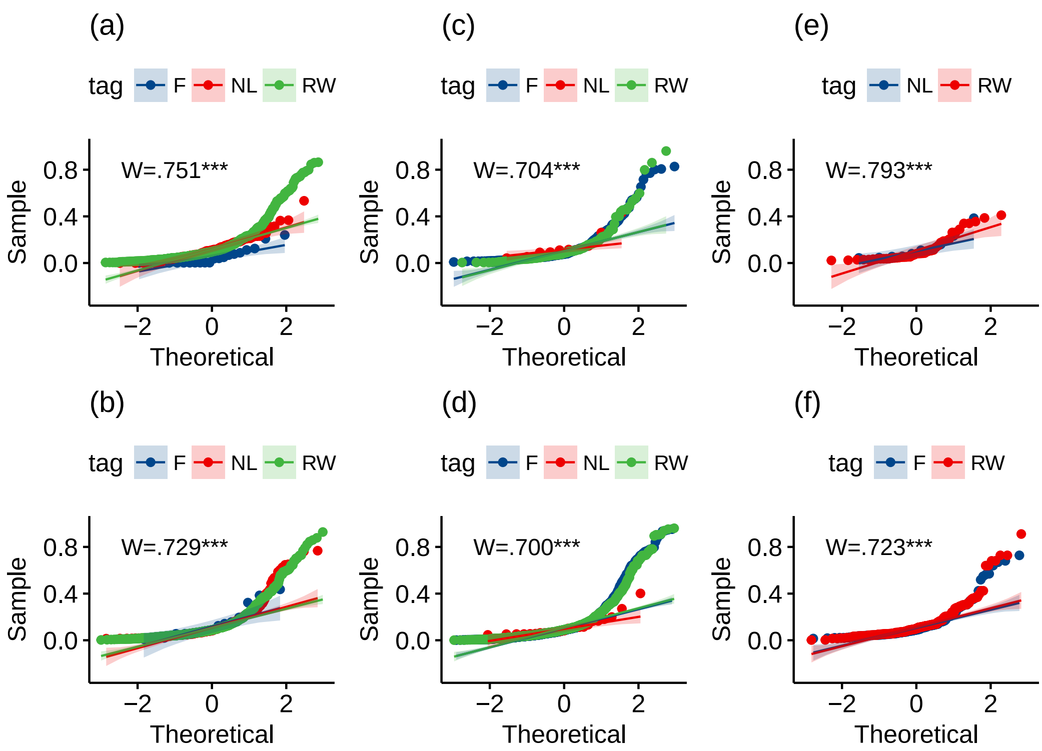

Figure 6.9: QQ plots and Shapiro-Wilk normality test values of the focus indices, extracted from classification of (FRWNL) natural scenes across the following time intervals: (ace) before official flood risk communication; (bdf) after official flood risk communication; (ab) during flood alerts; (cd) during flood warnings; (ef) during severe flood warnings.

Modeling time intervals were selected to answer the third big question of this thesis, specifically whether semantically unstable material has the potential to segment events according to their evolving severity. In the context of flooding this means that attention focus expressed via posted images should be significantly different during event stages, rather then before/after authoritative flood risk communication.

The null hypothesis was that attention focus on the spatio-temporal ensembles of images is represented by identical populations. To test the hypothesis, I applied the *kruskal.test* R function to compare the independent point data. Results presented in Table 6.1 illustrate the p-values for the three groups of lexemes. Here we can observe that semantically unstable lexical material (RW) is represented by the sets of images, which tend to vary compositionally in response to the event severity rather than to the authoritative flood warning communication. This behavior to the certain degree replicates the pattern of direct event descriptors (F), which demonstrates composition of *non-identical populations* extracted from the images posted during various types of warnings, but identical before and after risk communication. Interestingly, the benchmark lexemes (NL) demonstrated completely opposite trend, where event severity showed no observable influence on the compositional structures of the images posted during flood events on the floodplains, but where alerts and warnings led to significant difference in the composite focus of the natural scenes.

	Before/after warnings	**Across warning types**
F	1.24 NS	8.69 **
RW	1.97 NS	3.83 *
NL	4.63 **	0.38 NS

Table 6.1: Kruskal-Wallis H test (** = $p<0.05$; * = $p<0.1$).

6.5 Conclusion

The importance of this analysis lies in the fact that making use of social media can help us to substantially expand operational knowledge regarding the locations of the most vulnerable populations during hazardous events, as well as to make use of valuable local knowledge of how to efficiently manoeuvre using local landmarks and their semantic connectivity. These strategies generally align with risk perception studies, highlighting the importance of social insights for designing and evaluating risk communication programmes, as without detailed knowledge of behavioural *contexts* of targeted audiences, risk communication is unlikely to succeed.

The first set of informative findings here suggest that significant scope exists for exploration of use of the abstract language during extreme and risky events, which can be significantly enhanced with application of the multimodal techniques. This statement is also supported by the earlier research in social psychology about situational use of the abstract language [Smith and Trope, 2006; Wakslak et al., 2014] by the distinct groups, which can be defined as more "confident".

In this analysis, the structural dissimilarity of the associated with the neutral tags scenes across both sets (positive and risk-signalling), which decreases with event severity, is also indicative of the discriminatory potential for the severity of evolving hazards before authoritative risk communication takes place. Our statistical analysis confirms this, however, it can be advised that such analysis would benefit from testing across wider range of the hazard events or risk-related situations before the definite conclusion about the full potential of semantically drifted material for event segmentation on social media is made.

Chapter 7

Results and Discussion

Since, to the best of my knowledge, no study so far attempted to look into the concept of *event-driven* semantic drift, I needed to start off with the question of whether we can detect events with help of semantic drift on social media (i.e., whether words, which are ontologically close to the direct event descriptor, exhibit similar to it behaviour throughout duration of the event). A *frequency*-based [Kulkarni et al., 2015] approach was selected in order to demonstrate the quantitative co-fluctuations of the candidates for semantic drift ('river', 'water') with either direct event descriptor 'flood' or 'benchmark' lexemes ('nature', 'landscape'). This experiment confirmed that meaning drift of the words-candidates is indeed *transient* and the strength of relationships with the direct event descriptors appears to be higher several days before the peak of the event, thus indicating potential routes for exploring the *predictive* capacity of the semantically unstable lexical material on social media.

From the perspective of distributional semantics, what I proposed here can be regarded as an emerging *hybrid method design*, illustrating how to capture periodical word meaning broadenings and narrowings that form a basis for *transient semantic shifts*. While ongoing theoretical discussions in the field of semantic change are concerned with the uni-directional (irreversible) semantic changes playing a role in long-term language evolution, my method is mostly concerned with *short-term* and *reversible* changes that, according to some scholars are not lexical drifts at all, and could be seen as instances of semantic instability [Tredici et al., 2019].

Whist the first experiment provided some promising results, there are, nevertheless, some obvious limitations. For instance, targeting more numerous posts than direct event descriptor, I used only a very small number of the words-candidates for semantic drift, whilst ignoring less numerous *individually*, but potentially more numerous in combination. Incorporation of the *complete ontological networks* in this type of event analytics can provide promising avenues for future studies, which can critically evaluate and compare performance of the words-candidates. Secondly, whilst frequency-based approaches are generally acceptable by some corpus linguists [Hamilton et al., 2016b; Kulkarni et al., 2015], some others [Dubossarsky et al., 2017; Antoniak and Mimno, 2018] find it ambiguous, so the scope still remains for testing embedding-based approaches (word2vec, GloVe, PPMI, neural nets) [Pennington et al., 2014; Levy et al., 2015], whilst accounting for ontological similarity.

Once the tendency for *situational* semantic change was detected, I wanted to further verify how valuable those data signals are from the *spatial perspective* and whether they are capable of the *differentiating* flooding types, which can be specific for different areas due to their repetitive nature. For this purpose I used UK-wide spatial designations of the four main hydrometric networks, notably groundwater, precipitation, river flow and surface water stations, and I decided to test temporally aggregated (2004-2014) social media data points for the presence of spatially significant clusters. Initial sets of results (Global Moran's I demonstrated that there is potential for differentiating 'slow' floods (e.g., groundwater and precipitation (or pluvial) ones) from the 'fast' floods (surface water and river flow ones).

When looking at the structure of the local indicators of spatial autocorrelation (LISA), we can notice a slight segregation of results, where surface water ('SW') and precipitation ('Pr') scenarios constitute a group with the lowest trend of combined semantic drift (FRW), whilst river flow ('RF') and groundwater ('GW') form the group with the highest trend, although still smaller than in case when lexemes (R) and (W) are modelled individually (i.e., (FR(and (FW), respectively).

And returning back to our research question, I can confirm the results obtained during the global autocorrelation modeling and say that *surface water* and *river* floods are the ones, that provoke the highest volumes of semantically drifted material, presumably indicating more radical *shift of attention* from more environmentally neutral objects/topics. However, in the context of their global autocorrelation, the clusters they produce are much weaker with less prominent locations, which means that they are much less *localised* and much more widespread that *pluvial* and *groundwater* floods.

From the data selection perspective, the importance of accounting for single- or combined scenarios became more obvious when we look at the *structure* of *significant* spatial relationships between the modelled units. Fig 5.8 shows that combined models represent the highest potential to detect hot- (where event takes place) and cold spot (where event definitely does not) cluster areas, whereas single lexemes are only indicative of the processes occurring nearby.

This analysis provided some promising results for differentiation of the two main groups of flooding types, using existing infrastructure of the specialised sensors. As it became obvious during presentation of the results, I can suggest two research avenues for the follow-up studies.

The first one is linked to more critical selection of the *candidates* for semantic drift; Although ontologically approximated lexemes showed some promising results, there is nevertheless a scope for more systematic analysis of individual candidates, as well as for testing their combined performance. This type of analysis can provide some *useful* for environmental monitoring *socio-linguistic profiles* of the areas, where use of certain words or their combinations by the local residents can signal certain types of phenomena.

The second one is linked to more critical modelling approaches; Whilst Moran's I can be seen as an optimal model, the scope still remains to validate these results across other indices of spatial dependency [Fan and Myint, 2014]. There is also a scope for more critical, distributed analyses (e.g., 'time-space cubes' [TSC, 2019]), however, this is adviseable for the scenarios with more numerous candidates for semantic drift.

The idea that shift of attention (i.e., *contextual attention shift*) during hazard (or any other critical situation) conditions the *true meaning* behind the words people use was further followed up in the Chapter 6. From the practical perspective, I also wanted to verify how *event stages*, which are used by the Environment Agency and the Met Office in the UK, are reflected in the patterns of the collective crowd attention throughout flooding hazards. This idea was borrowed from the *market segmentation strategies*, based on behavioural profiles of the customers for more efficient targeting and adopted in the early 20s in the last century [Dickson and Ginter, 1987].

This set of results demonstrated that different types of crowdsourced lexical material with associated visual media related to the topic of environmental perception of risk have the potential to not only sense an approaching flooding hazard, but also indeed provide some insights into its *stages*, (ranging from the *least* to the *most severe*), thus providing the basis for *situational event segmentation strategies.* The main mechanisms behind such segmentations are linked to the active and passive ways in which people tend to act and react in conditions of uncertainty [Rotenberg, 2009], specifically a proactive search for visual cues in the surrounding environment before official warnings are issued [Tkachenko et al., 2016b]. Nevertheless, although some valuable patterns in data analysis started emerging, it was possible to conclude that this analysis will significantly benefit from testing on far more extensive datasets and potentially across much wider range of events.

Chapter 8

Conclusions

The main purpose of this thesis was to understand how to approach a specific natural hazard (i.e., flooding) from the *social event* perspective by using data, generated by event participants before, during and after the event(s) in order to assist to the design of the *socially adapted* flood risk management strategies. I used multimodal social media platform *Yahoo! Flickr* in order to propose three methods, aiming to answer currently problematic areas in flood risk monitoring, such as event *detection* [Wilkinson et al., 2015], *differentiation* [Wilkinson et al., 2014] and *segmentation* [Wang et al., 2016].

Social media has already been widely used in various events analytics studies [Alexander, 1991], including hazard monitoring [Earle, 2010; Acar and Muraki, 2011; Weng et al., 2011; Aggarwal and Abdelzaher, 2012; Al-Saggaf and Simmons, 2015; Preis et al., 2013; Peary et al., 2012] and mobility of the crowds during catastrophic events [Wang and Taylor, 2018]. Data selection for such studies is usually made on the basis of georeferenced keywords, which are direct event descriptors (e.g., 'flood', 'storm', 'avalanche'). As a consequence [I hypothesised], a lot of material, which can be potentially useful for event monitoring, is being omitted. To address this gap, I turned my attention to the concept of *semantic drift*, which is a well known category in distributional semantics and recently received a lot of attention from the web-focused corpus linguistics researchers [Hamilton et al., 2016b; Antoniak and Mimno, 2018; Dubossarsky et al., 2017; Kulkarni et al., 2015]. Whilst the main practical application of this concept is to understand diachronic dynamics of language, I decided to repurpose these recent findings (also known as 'Laws of Semantic Change' [Hamilton et al., 2016b]) to verify whether semantic drift also occurs around flood events and whether it generates additional data signals, which can be used to answer research questions of this thesis.

Following these theories, rather than standard embeddings-based approach of the corpus lingusitics, I manually selected lexemes-candidates for semantic drift, based on their *ontological* relationships [Fillmore, 1991]. This was done with the purpose to demonstrate behaviour of more numerous data signals around flood events, but also, more importantly, to develop a method for *corpus-, period- and timestep-independent* change of meanings. Following this principle, the latter would have suited analysis of both *event-driven* and *natural* semantic drifts across various corpora, however, in the scope of my analysis I concentrated on the single one, comprising several modalities. There is obviously a scope for the future studies to verify the behaviour of event-driven drifts across different corpora, for example, on Twitter and Facebook.

According to some emerging views in data science, at some point in the near time *theory-free* data-driven models will entirely supplant models that explicitly start from theory [Lyon, 2015]. From this perspective environmental and climate research look as interesting test cases for the suggested shift from process-based to so called "theory-free modelling" [Knüsel et al., 2019] as so far 'big data' typically has been applied to 'small problems' of the structured cases of repeated evaluation of predictions. Very recently, [Knüsel et al., 2019] have provided some critical overview towards the case of 'big' environmental analytics by pointing out intermediate categories between classical domain science and 'big data', thus implying that "*...Big-data elements can be useful for climate research beyond small problems if combined with more traditional approaches based on domain-specific knowledge...*". It can also be argued that where social media data is concerned as a 'big data' source, the tendencies towards its complementary modeling alongside existing sensors' signals [Restrepo-Estrada et al., 2018] can be also explained by the present scepticism due to inability to predict platform longevity or continuity of the API services or incapacity to find the ways to tackle its spatio-temporal inconsistencies [Restrepo-Estrada et al., 2018]. As a consequence, although behavioural event analytics is gaining an increasing attention amongst researchers and practitioners, natural hazard analytics is, nevertheless, still heavily grounded in techno-instrumental tradition of the physical sciences, where people are treated as passive agents rather than active event participants and decision makers.

Since event analytics is essentially concerned with pursuit of the main pragmatic aims of *detection*, *differentiation* and *segmentation* of phenomena, these tasks, whilst also representing the outstanding questions in flood risk management, have been repurposed in the context of this thesis for analysis of the hydrological risks from perspective of exposed populations, whose real-world experiences are captured by various modalities of the social media data (i.e., lexical, visual, geolocational).

I decided to use semantic drift as an emerging analytical tool whilst being primarily driven by the research interest to verify how it can help to answer some of the most pressing questions for flood risk management. However, I was also intrigued by the possibility to verify what can happen if we decide to develop methods, where traditional data and measurements lose their position of a *primary* information components and are used as a mere reference/background source. In such ways, I contemplated, we can see how far we can get when using social media as one of the main sources of information for risk communication and decision-making.

Bibliography

Events. https://plato.stanford.edu/entries/events/, 2002. Accessed: 2018-09-30.

Mapping Crime: Understanding Hot Spots. www.ojp.usdoj.gov/nij, 2005. Accessed: 2018-09-10.

Cumbria floods: Unpredictable but not unforeseeable. https://www.theguardian.com/commentisfree/2009/nov/24/cockermouth-cumbria-flood-hilary-benn, 2009. Accessed: 2018-04-14.

Ash sampling guidelines. https://www.bgs.ac.uk/research/volcanoes/AshCollection.html, 2011. Accessed: 2018-09-05.

Earthquake Information. http://earthquakes.bgs.ac.uk/questionnaire/EqQuestIntro.html, 2011. Accessed: 2018-09-05.

GeoExposures. https://www.bgs.ac.uk/citizenScience/geoexposures.html, 2011. Accessed: 2018-09-05.

GeoSocial. https://www.bgs.ac.uk/citizenScience/geosocial/home.html, 2011. Accessed: 2018-09-05.

EM-DAT: The International Disaster Database. https://www.emdat.be, 2011. Accessed: 2016-11-02.

iGeology App — Britain's rocks in your pocket. https://www.bgs.ac.uk/igeology/, 2011. Accessed: 2018-09-05.

mySoil App — Growing our knowledge. https://www.bgs.ac.uk/mySoil/, 2011. Accessed: 2018-09-05.

Flooding from groundwater: Practical advice to help you reduce the impact of flooding from groundwater. http://www.groundwateruk.org/downloads/EAGWFlooding.pdf, 2011. Accessed: 2016-07-05.

Report a Landslide. https://www.bgs.ac.uk/landslides/report.html, 2011. Accessed: 2018-09-05.

UK School Seismology Project. https://www.bgs.ac.uk/schoolseismology/schoolSeismology.cfc?method=viewLatestQuake, 2011. Accessed: 2018-09-05.

Filtering Tweets by Location. https://support.gnip.com/articles/filtering-twitter-data-by-location.html, 2013. Accessed: 2017-10-12.

word2vec. https://code.google.com/archive/p/word2vec/, 2013. Accessed: 2018-06-12.

UK storms: Mapping the floods. https://www.bbc.co.uk/news/uk-26127121, 2014. Accessed: 2019-01-05.

What is the difference between a heat map and a hot spot map? https://www.gislounge.com/difference-heat-map-hot-spot-map/, 2014. Accessed: 2018-09-10.

Groundwater flooding in the UK: January 2014. https://www.bgs.ac.uk/research/highlights/2014/groundwaterFloodingJan.html, 2014. Accessed: 2015-01-26.

Places Database. http://places.csail.mit.edu/index.html, 2014. Accessed: 2016-05-07.

Language detectives make the web less anonymous. https://www.cbsnews.com/news/forensic-linguistics-makes-web-less-anonymous/, 2015. Accessed: 2019-01-05.

Peter Moore: The Weather Experiment. http://peter-moore.co.uk/books/the-weather-experiment, 2015. Accessed: 2017-06-15.

The National Archives. https://webarchive.nationalarchives.gov.uk/search/, 2015. Accessed: 2015-06-14.

Word Mapper. https://isogloss.shinyapps.io/isogloss/, 2015. Accessed: 2018-09-05.

WordNet: A Lexical Database for English. https://wordnet.princeton.edu, 2015. Accessed: 2018-06-12.

anelachan/sentimentanalysis. https://github.com/anelachan/sentimentanalysis, 2016. Accessed: 2018-07-14.

Geolexicography. http://languagelog.ldc.upenn.edu/nll/?p=23780, 2016. Accessed: 2018-11-10.

Why next generation social media is visual. https://www.marketingtechnews.net/news/2016/jun/23/why-next-generation-social-media-visual/, 2016. Accessed: 2018-05-14.

Build on flood plains despite the risks, say UK government advisers. https://www.theguardian.com/environment/2016/jan/27/homes-and-companies-should-be-built-on-flood-plains-despite-risks-says-panel, 2016. Accessed: 2018-04-14.

Weather Observations Website. https://wow.metoffice.gov.uk, 2018. Accessed: 2018-09-05.

Burwell Laboratory of Memory and Attention. https://www.brown.edu/research/labs/burwell/, 2019. Accessed: 2018-07-07.

Gestalt Laws of Perceptual Organization. https://www.verywellmind.com/gestalt-laws-of-perceptual-organization-2795835, 2019. Accessed: 2018-12-14.

Defra Data Services Platform. https://environment.data.gov.uk, 2019. Accessed: 2018-09-05.

Flood Forecasting Centre. http://www.ffc-environment-agency.metoffice.gov.uk, 2019. Accessed: 2018-09-05.

What are the two types of floods? https://www.usgs.gov/faqs/what-are-two-types-floods?qt-news_science_products=0#qt-news_science_products, 2019. Accessed: 2018-09-05.

5-day flood risk for England and Wales. https://flood-warning-information.service.gov.uk/5-day-flood-risk, 2019. Accessed: 2018-09-05.

GeoDa: An Introduction to Spatial Data Analysis. https://geodacenter.github.io, 2019. Accessed: 2017-08-14.

Institute for Watershed Science. https://www.trentu.ca/iws/, 2019. Accessed: 2018-09-05.

Weather Research and Forecasting Model. https://www.mmm.ucar.edu/weather-research-and-forecasting-model, 2019. Accessed: 2018-09-05.

Find a forecast. https://www.metoffice.gov.uk, 2019. Accessed: 2018-09-05.

Neighbourhood Alert: Informing and Listening to Communities. https://www.neighbourhoodalert.co.uk, 2019. Accessed: 2018-09-05.

NWS High Water Level Terminology. https://www.weather.gov/aprfc/terminology, 2019. Accessed: 2018-11-13.

A Beginner's Guide To Grid References. https://getoutside.ordnancesurvey.co.uk/guides/beginners-guide-to-grid-references/, 2019. Accessed: 2017-04-04.

Past Weather Events. https://www.metoffice.gov.uk/weather/learn-about/past-uk-weather-events, 2019. Accessed: 2018-09-05.

How Creating a Space Time Cube works. https://pro.arcgis.com/en/pro-app/tool-reference/space-time-pattern-mining/learnmorecreatecube.htm, 2019. Accessed: 2018-12-12.

Unified Model. https://www.metoffice.gov.uk/research/approach/modelling-systems/unified-model/index, 2019. Accessed: 2018-09-05.

A Acar and Y Muraki. Twitter for crisis communication: Lessons learned from Japan's Tsunami disaster. *International Journal of Web Based Communities*, 7 (3):392–402, 2011.

JCJH Aerts, WJ Botzen, KC Clarke, SL Cutter, JW Hall, B Merz, E Michel-Kerjan, J Mysiak, S Surminski, and H Kunreuther. Integrating human behaviour dynamics into flood disaster risk assessment. *Nature Climate Change*, 8:193–199, 2018.

CC Aggarwal and T Abdelzaher. Social Sensing. *In: Aggarwal C. (Eds.) Managing and Mining Sensor Data. Springer, Boston, MA*, pages 237–297, 2012.

A AghaKouchak, LS Huning, F Chiang, M Sadegh, F Vahedifard, O Mazdiyasni, H Moftakhari, and I Mallakpour. How do natural hazards cascade to cause disasters? *Neuron*, 561:458–460, 2018.

Y Al-Saggaf and P Simmons. Social media in Saudi Arabia: Exploring its use during two natural disasters. *Technological Forecasting and Social Change*, 95(3):3–15, 2015.

D Alexander. Natural disasters: A framework for research and teaching. *Disasters*, 15(3):209–226, 1991.

T Allison, A Puce, and G McCarthy. Social perception from visual cues: Role of the STS region. *Nature*, 439(7075):462–465, 2006.

L Anselin, I Syabri, and O Smirnov. Visualizing multivariate spatial correlation with dynamically linked windows. *New Tools for Spatial Data Analysis: Proceedings of the Specialist Meeting; Santa Barbara*, 2002.

L Anselin, J Le Gallo, and H Jayet. Spatial panel econometrics. *The Econometrics of Panel Data, Fundamentals and Recent Developments in Theory and Practice*, pages 625–660, 2008.

M Antoniak and D Mimno. Evaluating the stability of embedding-based word similarities. *Transactions of the Association for Computational Linguistics*, 6:107–119, 2018.

JJ Appleton. *The Experience of Landscape*. London: Wiley., 1975.

A Arntz, LW Hawke, L Bamelis, P Spinhoven, and ML Molendijk. Changes in natural language use as an indicator of psychotherapeutic change in personality disorders. *Behavioural Research and Therapy*, 50(3):191–202, 2012.

BF Atwater, S Musumi-Rokkaku, K Satake, Y Tsuji, K Ueda, and DK Yamaguchi. The Orphan Tsunami of 1700: Japanese clues to a parent earthquake in North America. U.S. Geological Survey Professional Paper 1707:135, 2005.

A Bakarov. A survey of word embeddings evaluation methods. *arXiv preprint arXiv:1801.09536*, 2018.

GD Baldassarre, M Kooy, JS Kemerink, and L Brandimarte. Towards understanding the dynamic behaviour of floodplains as human-water systems. *Hydrol Earth Syst Sci*, 17:3235–3244, 2013.

S Banerjee and T Pedersen. Extended Gloss Overlaps as a Measure of Semantic Relatedness. *Proceedings of the International Joint Conference on Artificial Intelligence*, 2003.

LW Barsalou. Perceptual Symbol Systems. *Behavioural and Brain Sciences*, 22(4): 610–660, 1999.

A Bashawria, S Garritya, and K Moodleya. An overview of the design of disaster relief shelters. *Procedia Economics and Finance*, 18:924–931, 2014.

F Bavaud. Models for spatial weights: A systematic look. *Geographical Analysis*, 30:153–171, 1998.

L Bengtsson, X Lu, A Thorson, R Garfield, and J von Schreeb. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti. *PLoS Medicine*, 8(8): e1001083, 2011.

NA Bernstein. *The co-ordination and regulation of movements*. Oxford: Pergamon Press., 1967.

J Betjeman. *Slick but not Streamlined*. New York: Doubleday and Co., 1947.

K Beven. Towards integrated environmental models of everywhere: Uncertainty, data and modeling as a learning process. *Hydrol Earth Syst Sci*, 11:460–467, 2007.

S Birkholz, M Muro, P Jeffrey, and HM Smith. Rethinking the relationship between flood risk perception and flood management. *Science of the Total Environment*, 478:12–20, 2014.

P Blair and W Buytaert. Socio-hydrological modelling: A review asking 'why, what and how?'. *Hydrol Earth Syst Sci*, 20:443–478, 2016.

VV Bochkarev, AV Shevlyakova, and VD Solovyev. The average word length dynamics as an indicator of cultural changes in society. *Social Evolution and History*, 14(2):153–175, 2015.

B Boodram, AL Hotton, L Shekhtman, A Gutfraind, and H Dahari. High-Risk Geographic Mobility Patterns among Young Urban and Suburban Persons who Inject Drugs and their Injection Network Members. *Journal of Urban Health*, 95 (1):71–82, 2018.

MNK Boulos, B Resch, DN Crowley, JG Breslin, G Sohn, and R Burtner. Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: Trends, ogc standards and application examples. *International Journal of Health Geographics*, 10(1):1–29, 2011.

TGR Bower. The determinants of perceptual unity in infancy. *Psychometric Science*, 3:323–324, 1965.

M Bréal. *Essai de Sémantique: Science des Significations.* Paris: Hachette., 1897.

D Brockmann, L Hufnagel, and T Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, 2006.

G Brogueira, F Batista, and JP Carvalho. Using geolocated tweets for characterization of Twitter in Portugal and the Portuguese administrative regions. *Social Network Analysis and Mining*, 6(37), 2016.

JD Brown and SL Damery. Managing flood risk in the UK: Towards an integration of social and technical perspectives. *Transactions of the Institute of British Geographers*, 27:412–426, 2002.

C Brugman and G Lakoff. Cognitive topology and lexical networks. *In: GW Cottrell, S Small and MK Tannenhause (Eds.) Lexical Ambiguity Resolution: Perspectives from Psycholinguistics, Neuropsychology and Artificial Intelligence. San Mateo, CA: Morgan Kaufman Publishers*, 1988.

TE Bunting and L Guelke. Behavioral and Perception Geography: A Critical Appraisal. *Annals of the Association of American Geographers*, 69(3):448–462, 1979.

I Burton and RW Kates. The perception of natural hazards in resource management. *Natural Resources Journal*, 3(3):412–441, 1964.

J Chakraborty, GA Tobin, and BE Montz. Population evacuation: Assessing spatial variability in geophysical risk and social vulnerability to natural hazards. *Natural Hazards Review*, 6(1):23–33, 2005.

E Chan, O Baumann, MA Bellgrove, and JB Mattingley. From objects to landmarks: The function of visual location information in spatial navigation. *Frontiers in Psychology*, 3:304, 2012.

A Chioda. *Monitoring urban mobility with social networks.* Master graduation thesis submitted to Politecnico di Milano., 2014.

Y Chun and DA Griffith. *Spatial Statistics and Geostatistics.* SAGE Publications, 2013.

AD Cliff and JK Ord. *The Problem of Spatial Autocorrelation.* London: Pion., 1969.

MA Cohen, GA Alvarez, and K Nakayama. Natural-scene perception requires attention. *Psychological Science*, 22(9):1165–1172, 2011.

R Cohen, S Havlin, and D Ben-Avraham. *Structural properties of scale free networks. Handbook of graphs and networks.* Wiley-VCH., 2002.

V Cologna, RH Bark, and J Paavola. Flood risk perceptions and the UK media: Moving beyond 'once in a lifetime' to 'Be Prepared' reporting. *Climate Risk Management*, 17:1–10, 2017.

National Research Council. *Geotargeted Alerts and Warnings: Report of a Workshop on Current Knowledge and Research Gaps.* Washington, DC: The National Academies Press., 2013.

A Crooks, A Croitoru, A Stefanidis, and J Radzikowski. Earthquake: Twitter as a distributed sensor system. *Transactions in GIS*, 17(1):124–147, 2013.

RJ Dawson, R Peppe, and M Wang. An agent-based model for risk-based flood incident management. *Natural Hazards*, 59:167–189, 2011.

S Deerwester, ST Dumais, G Furnas, T Landauer, and R Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41:391–407, 1990.

L Degrossi, J Albuquerque, C Restrepo-Estrada, A Mobasheri, and A Zipf. Exploring the geographical context for quality assessment of VGI in flood management domain. *Proceedings of 14th ISCRAM conference*, 2017.

J Diaz-Nieto, DN Lerner, A Saul, and J Blanksby. GIS Water-Balance Approach to Support Surface Water Flood-Risk Management. *Journal of Hydrologic Engineering*, 17:55–67, 2012.

PR Dickson and JL Ginter. Market Segmentation, Product Differentiation, and Marketing Strategy. *Journal of Marketing*, 51(2):1, 1987.

D Divjak, E Dabrowska, and A Arppe. Machine meets man: Evaluating the psychological reality of corpus-based probabilistic models. *Cognitive Lingusitics*, 27 (1):1–33, 2016.

E Du, S Rivera, X Cai, L Myers, A Ernest, and B Minsker. Impacts of human behavioural heterogeneity on the benefits of probabilistic flood warnings: An agent-based modeling framework. *Journal of the American Water Resources Association*, 53(2):316–332, 2016.

A Dubey. Deep learning the city: Quantifying urban perception at a global scale. *European Conference on Computer Vision*, 2016.

H Dubossarsky, D Weinshall, and E Grossman. Outta Control: Laws of Semantic Change and Inherent Biases in Word Representation Models. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1136–1145, 2017.

J Durbin. Weak convergence of the sample distribution function when parameters are estimated. *The Annals of Statistics*, 1(2):279–290, 1973.

P Earle. Earthquake Twitter. *Nature Geoscience*, 3:221–222, 2010.

J Eisenstein. Unsupervised learning for lexicon-based classification. *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 2017.

AD Ekström, AEGF Arnold, and G Iaria. A critical review of the allocentric spatial representation and its neural underpinnings: Toward a network-based perspective. *Frontiers in Human Neuroscience*, 8:803, 2014.

D Elkind and L Scott. Studies in Perceptual Development I: The Decentering of Perception. *Child Development*, 33:619, 1962.

O Etzion and P Niblett. *Event Processing in Action.* Manning, Greenwich., 2011.

C Fan and S Myint. A comparison of spatial autocorrelation indices and landscape metrics in measuring urban landscape fragmentation. *Landscape and Urban Planning*, 121:117–128, 2014.

G Fattoruso, A Agresta, S De Vito, A Buonanno, M Molinara, C Marrocco, F Tortorella, and G Di Francia. Online anomaly detection on rain gauge networks for robust alerting services to citizens at risk from flooding. *In: Gervasi O. et al. (Eds.) Computational Science and Its Applications*, 10406:427–442, 2017.

T Filatova. Empirical agent-based land market: Integrating adaptive economic behaviour in urban land-use models. *Computers, Environment and Urban Systems*, 54:397–413, 2015.

CJ Fillmore. 'Corpus linguistics' vs. 'Computer-aided armchair linguistics'. *Directions in Corpus Linguistics: Proceedings from a 1991 Nobel Symposium on Corpus Linguistics*, pages 35–66, 1991.

MM Fischer and A Getis. *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications.* Springer-Verlag Berlin Heidelberg., 2010.

K Freberg, K Saling, KG Vidoloff, and G Eosco. Using value modeling to evaluate social media messages: The case of Hurricane Irene. *Public Relations Review*, 39 (3):185–192, 2013.

SM Freundschuh and MJ Egenhofer. Human conceptions of spaces: Implications for Geographic Information Systems. *Transactions in GIS*, 2(4):361–375, 1997.

T Fujiyama, C Matsui, and A Takemura. A power-law growth and decay model with autocorrelation for posting data to social networking services. *PLoS ONE*, 11:e0160592, 2016.

P Gaenssler and W Stute. Empirical processes: A survey of results for independent and identically distributed random variables. *The Annals of Probability*, 7(2): 193–243, 1979.

JAE Gaitan, S van de Giesen, and NC ten Veldhuis. Can urban pluvial flooding be predicted by open spatial data and weather data? *Environmental Modelling and Software*, 8:156–171, 2016.

D Garcia, A Garas, and F Schweitzer. Positive words carry less information than negative words. *EPJ Data Science*, 1(3):1–12, 2012.

P Gärdenfors. *Conceptual Spaces: The Geometry of Thought.* MIT Press., 2000.

P Gärdenfors. Cognitive semantics and image schemas with embodied forces. *In: JM Krois, M Rosengren, A Steidele and D Westerkamp (Eds.) Embodiment in Cognition and Culture. John Benjamins Publishing Company*, pages 57–76, 2007.

D Geeraerts. *Cognitive linguistics and the history of philosophical epistemology. In R Geiger and B Rudzka-Ostyn, eds., Conceptualizations and mental processing in language.* Berlin: Mouton de Gruyter., 1993.

A Getis. Reflections on spatial autocorrelation. *Regional Science and Urban Economics*, 37:491–496, 2007.

A Getis. A history of the concept of spatial autocorrelation: A geographer's perspective. *Geographycal Analysis*, 40:297–309, 2008.

A Getis and JK Ord. The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24:189–206, 1992.

RW Gibbs and HL Colston. The cognitive psychological reality of image schemas and their transformations. *Cognitive Linguistics*, 6(4):347–378, 1995.

JJ Gibson. *The ecological approach to the visual perception.* Boston: Houghton Mifflin., 1979.

G Gilquin. Taking a new look at lexical networks. *Lexis*, 1:23–39, 2008.

CJ Glacken. *Traces on the Rhodian Shore: Nature and Culture in Western Thought from Ancient Times to the end of the Eighteenth Century.* Berkeley: University of California Press., 1967.

VG Glushkov. *Geographical and hydrological method.* Izh. GGI. No.57-58, 1933.

D Glynn and J Robinson. *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy.* John Benjamins Publishing Company., 2014.

JR Gold. *An Introduction to Behavioural Geography.* Oxford University Press, 1980.

D Gonzales, L Kraus, J Osburg, SR Shelton, and D Woods. *Geo-targeting Performance of Wireless Emergency Alerts in Imminent Threat Scenarios.* Published in: Geo-Targeting Performance of Wireless Emergency Alerts in Imminent Threat Scenarios (Washington, D.C.: U.S. Department of Homeland Security (DHS) Science and Technology (SandT) Directorate, June 2016)., 2016.

MF Goodchild. Geographic Information Systems and Science: Today and Tomorrow. *Annals of Geographical Information Science*, 15(1):3–9, 2009.

IC Goulter and NM Myska. The human component in flood warning and flood response system. *GeoJournal*, 15(3):297–305, 1987.

J Grieve. A regional analysis of contraction rate in written Standard American English. *International Journal of Corpus Linguistics*, 16:514–546, 2011.

J Grieve. A statistical comparison of regional phonetic and lexical variation in American English. *Literary and Linguistic Computing*, 28:82–107, 2013.

J Grieve, D Speelman, and D Geeraerts. A statistical method for the identification and aggregation of regional linguistic variation. *Language Variation and Change*, 23(2):193–221, 2011.

J Grieve, A Nini, and D Guo. Analysing lexical emergence in American English online. *English Language and Linguistics*, 21:99–127, 2017.

DW Hamacher and RP Norris. Australian aboriginal geomythology: Eyewitness accounts of cosmic impacts? *Archaeoastronomy: the Journal of Astronomy in Culture*, 22:62–95, 2010.

T Hamaguchi, H Oiwa, M Shimbo, and Y Matsumoto. Knowledge Transfer for Out-of-Knowledge-Base Entities: A Graph Neural Network Approach. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pages 1802–1808, 2017.

WL Hamilton, K Clark, J Leskovec, and D Jurafsky. Inducing domain-specific sentiment lexicons from unlabeled corpora. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 595–605, 2016a.

WL Hamilton, J Leskovec, and D Jurafsky. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, page 1489–1501, 2016b.

J Handmer and B Proudley. Communicating uncertainty via probabilities: The case of weather forecasts. *Environmental Hazards*, 7(2):79–87, 2007.

R Harris. The handbook of spatial statistics. *International Journal of Geographical Information Science*, 25:333–334, 2011.

L Herranz, S Jiang, and X Li. Scene recognition with CNNs: Objects, scales and dataset bias. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 571–579, 2018.

R Heuser. Word vectors in the eighteenth- century. *In IPAM workshop: Cultural Analytics*, 2016.

G Higginbottom. Sampling issues in qualitative research. *Nurse Researcher*, 12: 7–19, 2004.

J Hilferty. Mothers, lies, and bachelors: A brief reply to Wierzbicka (1990). *Word*, 48:51–59, 2015.

D Hilfinger, C Barrington, and E Lacy. Latino social network dynamics and the Hurricane Katrina disaster. *Disasters*, 36:101–121, 2011.

JB Hirsh and JB Peterson. Personality and language use in self-narratives. *Journal of Research in Personality*, 43(3):524–527, 2009.

G Hirst and D St-Onge. Lexical chains as representation of context for the detection and correction malapropisms. *WordNet: An Electronic Lexical Database*, 49:305–332, 1998.

K Holmqvist. *Implementing Cognitive Semantics.* PhD thesis, Lund University, 1993.

A Holzkämper, V Kumar, B Surridge, A Paetzold, and DN Lerner. Bringing diverse knowledge sources together - A meta-model for supporting integrated catchment management. *Journal of Environmental Management*, 96(1):116–127, 2012.

X Huang, C Wang, and Z Li. Reconstructing flood inundation probability by enhancing near real-time imagery with real-time gauges and tweets. *IEEE Transactions on Geoscience and Remote Sensing*, 56:4691–4701, 2018.

NE Humphries, N Queiroz, JRM Dyer, NG Pade, MK Musyl, KM Schaefer, DW Fuller, JM Brunnschweiler, TK Doyle, JDR Houghton, GC Hays, CS Jones, LR Noble, VJ Wearmouth, EJ Southall, and DW Sims. Environmental context explains Lévy and Brownian movement patterns of marine predators. *Nature*, 465:1066–1069, 2010.

ET Huurne and J Gutteling. Information needs and risk perception as predictors of risk information seeking. *Journal of Risk Research*, 11:847–862, 2008.

R Iliev, J Hoover, M Dehghani, and R Axelrod. Linguistic positivity in historical texts reflects dynamic environmental and psychological factors. *Proceedings of the National Academy of Sciences*, 113(49):7871–7879, 2016.

W Ittelson. *Environment Perception and Contemporary Perceptual Theory.* In W. H. Ittelson (Ed.), Environment and Cognition (pp. 141-154). New York: Seminar., 1973.

WH Ittelson. *Environment Perception and Contemporary Perceptual Theory.* New York: Holt, Rinehart and Winston., 1976.

P Jaccard. The distribution of the flora in the Alpine zone. *New Phytologist*, 11: 37–50, 1912.

B Jager and AA Cleland. Connecting the research fields of lexical ambiguity and figures of speech: Polysemy effects for conventional metaphors and metonyms. *The Mental Lexicon*, 10(1):133–151, 2015.

SZ Janif, PD Nunn, P Geraghty, W Aalbersberg, FR Thomas, and M Camailakeba. Value of traditional oral narratives in building climate-change resilience: Insights from rural communities in Fiji. *Ecology and Society*, 21(2):7, 2016.

P Jansen-Osmann and P Fuchs. Wayfinding behavior and spatial knowledge of adults and children in a virtual environment: The role of landmarks. *Experimental Psychology*, 53(3):171–181, 2006.

P Javelle, J Demargne, D Defrance, J Pansu, and P Arnaud. Evaluating flash-flood warnings at ungauged locations using post-event surveys: A case study with the AIGA warning system. *Hydrological Sciences Journal*, 59(7):1390–1402, 2014.

JJ Jiang and DW Conrath. Semantic similarity based on corpus statistics and lexical taxonomy. *ROCLING X*, page 9008, 1997.

M Johnson. *The body in the mind: The bodily basis of meaning, imagination and reason.* Chicago: University of Chicago Press., 1987.

NA Kacinik and C Chiarello. Understanding metaphors: Is the right hemisphere uniquely involved? *Brain and Language*, 100(2):188–207, 2007.

I Karamouzas, B Skinner, and JG Stephen. Universal Power Law Governing Pedestrian Interactions. *Physical Review Letters*, 113:238701, 2014.

MA Kaufhold and C Reuter. The self-organization of digital volunteers across social media: The case of the 2013 European floods in Germany. *Journal of Homeland Security and Emergency Management*, 13(1):137–166, 2016.

M Kavouras, M Kokla, and E Tomai. Comparing categories among geographic ontologies. *Computers and Geosciences*, 31(2):145–154, 2005.

RPC Kessels, A van Doormaal, and G Janzen. Landmark Recognition in Alzheimer's Dementia: Spared Implicit Memory for Objects Relevant for Navigation. *PLoS ONE*, 6(4):e18611, 2011.

Y Kim, Yi-I Chiu, K Hanaki, D Hegde, and S Petrov. Temporal analysis of language through neural language models. *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, 2014.

E Kittay and A Lehrer. Semantic fields and the structure of metaphor. *Studies in Language*, 5(1):31–63, 1981.

J Kleinberg. Navigation in a Small World. *Nature*, 406:845, 2000.

E Klepousniotou and RS Baum. Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics*, 20:1–24, 2007.

E Klepousniotou, D Titone, and C Romero. Making sense of word senses: The comprehension of polysemy depends on sense overlap. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6):1534–1543, 2008.

B Knüsel, M Zumwald, C Baumberger, G Hirsch Hadorn, EM Fischer, DN Bresch, and R Knutti. Applying big data beyond small problems in climate research. *Nature Climate Change*, 9:196–202, 2019.

HW Krohne. The concept of coping modes: Relating cognitive person variables to actual coping behaviour. *Advances in Behaviour Research and Therapy*, 11: 235–249, 1989.

B Kuipers. Modeling spatial knowledge. *Cognitive Science*, 2(2):129–153, 1978.

V Kulkarni, R Al-Rfou, B Perozzi, and S Skiena. Statistically Significant Detection of Linguistic Change. *Proceedings of the 24th International Conference on World Wide Web*, pages 625–635, 2015.

ZW Kundziewicz and L Somlyony. Climatic change impact on water resources in a systems perspective. *Water Resources Management*, 11:407–435, 1997.

W Labov. *Field methods of the project on linguistic change and variation. In J Baugh and J Sherzer (eds.) Language in Use: Readings in Sociolinguistics.* Upper Saddle River, New Jersey: Prentice Hall., 1984.

G Lakoff. *Women, fire and dangerous things.* Chicago: University of Chicago Press., 1987.

G Lakoff and M Johnson. *Metaphors We Live By.* Chicago: University of Chicago Press., 1980.

RW Langacker. *Foundations of cognitive grammar, vol.1: Theoretical prerequisites.* Stanford: Stanford University Press., 1987.

C Leacock and M Chodorow. Combining local context and WordNet similarity for word sense identification. *WordNet: An Electronic Lexical Database*, 49:265–283, 1998.

O Levy, Y Goldberg, and I Dagan. Improving distributional similarity with lessons learned from word embeddings. *Transactions of the Association for Computational Linguistics*, 3:211–225, 2015.

D Lin. An information-theoretic definition of similarity. *Proceedings of the International Conference on Machine Learning*, 15:296–304, 1998.

S Lopez-Aparicio, M Guevara, P Thunis, K Cuvelier, and L Tarrasón. Assessment of discrepancies between bottom-up and regional emission inventories in Norwegian urban areas. *AE*, 154:285–296, 2017.

D Lowenthal. The American Scene. *Geographical Review*, 58:61–88, 1964.

JG Lu, JJ Lee, F Gino, and AD Galinsky. Polluted morality: Air pollution predicts criminal activity and unethical behavior. *Psychological Science*, 29(3):340–355, 2018.

D Luckham. *The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems*. Addison-Wesley Longman Publishing Co., 2002.

A Lyon. *Data*. Published in: The Oxford Handbook of the Philosophy of Science (ed. Humphreys, P.) 738–758 (Oxford Univ. Press, Oxford, 2015)., 2015.

AM MacEachren. Visualizing Uncertain Information. *Cartographic Perspective*, (13):10–19, 1992.

EB Mackay, ME Wilkinson, CJA MacLeod, KJ Beven, BJ Percy, MG Macklin, PF Quinn, M Stutter, and PM Haygarth. Digital catchment observatories: A platform for engagement and knowledge exchange between catchment scientists, policy makers, and local communities. *Water Resources Research*, 51(6):4815–4822, 2015.

HA Mallot and S Gillner. Route navigating without place recognition: What is recognised in recognition-triggered responses? *Perception*, 29(1):43–55, 2000.

FW Mast and L Jäncke. *Spatial Processing in Navigation, Imagery and Perception*. Springer Science and Business Media, 2007.

JA Mawdsley, DB Ball, and AJ Adeloye. *Evaluating a streamgauge network: A case study of I. F. Karasev (1968) On the principles of hydrological network design and prospective development*. Trudy GGI, No 164, 3-36., 1990.

BG McAdoo and L Paravisini-Gebert. Not the Earthquake's fault. *Nature Geoscience*, 4, 2011.

T Meilinger, M Knauff, and HH üBlthoff. Working Memory in Wayfinding—A Dual Task Experiment in a Virtual City. *Cognitive Science*, 32, 2010.

AG Miller, R Beckwith, C Fellbaum, D Gross, and JK Miller. WordNet: An on-line lexical database. *Communications of the ACM*, 38, 2008.

F Miranda, H Doraiswamy, M Lage, K Zhao, B Goncalves, L Wilson, M Hsieh, and C Silva. Urban Pulse: Capturing the Rhythm of Cities. *IEEE Transactions on Visualization and Computer Graphics*, 23:791–800, 2016.

A Mishra and P Coulibaly. Developments in hydrometric network design: A review. *Reviews of Geophysics*, 47(2):1–24, 2009.

HS Moat, T Preis, CY Olivola, C Liu, and N Chater. Using big data to predict collective behavior in the real world. *Behavioural and Brain Sciences*, 37(1):92–93, 2014.

PAP Moran. Notes on continuous stochastic phenomena. *Biometrika*, 37:17–23, 1950.

HA Morrow-Jone and CR Morrow-Jone. Mobility due to natural disaster: Theoretical considerations and preliminary analyses. *Disasters*, 15(2):126–132, 1991.

MM Murray and MT Wallace. *The neural bases of multisensory processes.* CRC Press, 2011.

R Navigli. Word sense disambiguation: A survey. *ACM Computing Surveys*, 41, 2009.

J Nemec and AJ Askew. *Mean and variance in network design philosophies. Integrated design of hydrological networks.* Budapest, Hungary: IAHS., 1986.

G Newman, A Wiggins, A Crall, E Graham, S Newman, and K Crowston. The future of citizen science: Emerging technologies and shifting paradigms. *Frontiers in Ecology and the Environment*, 10(6):298–304, 2012.

J Newman. From polysemy to semantic change: Towards a typology of lexical semantic associations. *Language*, 86(2):473–475, 2010.

D Nguyen. *Text as social and cultural data: A computational perspective on variation in text.* PhD thesis, University of Twente., 2017.

S Niederer. *Networked images: Visual methodologies for the digital age.* Amsterdam: Hogeschool van Amsterdam., 2018.

J Oliver. The significance of natural hazards in a developing area: A case study from North Queensland. *Geography*, 60:99–110, 1975.

A Páez and D Scott. Spatial statistics for urban analysis: A review of techniques with examples. *GeoJournal*, 61:53–67, 2005.

X Pan, CS Han, K Dauber, and KH Law. A multi-agent based framework for the simulation of human and social behaviors during emergency evacuations. *AI and Society*, 22(2):113–132, 2007.

E Panofsky. *Meaning in the Visual Arts.* Harmondsworth: Penguin., 1970.

E Pardo-Iguzquiza. Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. *Journal of Hydrology*, 210:206–220, 1998.

DJ Parker, SJ Priest, and SM Tapsell SM. Understanding and enhancing the public's behavioural response to flood warning information. *Meteorological Applications*, 16(1):103–114, 2009.

BDM Peary, R Shaw, and Y Takeuchi. Utilization of social media in the East Japan Earthquake and Tsunami and its effectiveness. *Journal of Natural Disaster Science*, 34(1):3–18, 2012.

JW Pennebaker. Putting stress into words: Health, linguistic, and therapeutic implications. *Behavioural Research and Therapy*, 31(6):539–548, 1993.

J Pennington, R Socher, and C Manning. GloVe: Global Vectors for Word Representation. *Empirical Methods in Natural Language Processing*, 14, 2014.

DR Petrolia, CE Landry, and KH Coble. Risk preferences, risk perceptions, and flood insurance. *Landscape Economics*, 89:227–245, 2013.

B Poucet. Spatial cognitive maps in animals: new hypotheses on their structure and neural mechanisms. *Psychol Rev*, 100(2):163–182, 1993.

T Preis, HS Moat, SR Bishop, P Treleaven, and HE Stanley. Quantifying the digital traces of Hurricane Sandy on Flickr. *Nature Scientific Reports*, 3:3141, 2013.

F Pulvermüller. Brain reflections of words and their meaning. *Trends in Cognitive Sciences*, 5:517–524, 2001.

RS Pyrce. *Hydrological low flow indices and their uses. WSC Report No.04-2004.* Watershed Science Centre, Peterborough, Ontario., 2004.

D Quercia. The Pursuit of Urban Happiness. *Proceedings of the 23rd International Conference on World Wide Web*, pages 611–612, 2014.

AF Rainbird. Precipitation-basic principles of network design. *Symposium on Design of Hydrological Networks, WMO IASH Pub*, 67:19–30, 1967.

EB Reiter and GS Sripada. Human variation and lexical choice. *Computational Lingusitics*, 28:545–553, 2002.

O Renn. *Risk Governance. Coping with Uncertainty in a Complex World.* London: Earthscan., 2008.

P Resnik. Using information content to evaluate semantic similarity in a taxonomy. *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 1*, 95:448–453, 1970.

C Restrepo-Estrada, S Camargo de Andrade, N Abe, MC Fava, EM Mendiondo, and J Porto de Albuquerque. Geo-social media as a proxy for hydrometeorological data for streamflow estimation and to improve flood monitoring. *Computers and Geosciences*, 111:148–158, 2018.

D Riemer, L Stojanovic, and N Stojanovic. Using complex event processing for semantic requests in real-time social media monitoring. *AAAI Technical Report WS-12-02 Real-Time Analysis and Mining of Social Streams*, 2012.

H Riesch and C Potter. Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions. *Public Understanding of Science*, 23(1):107–120, 2013.

S Robert. Words and their meanings: Principles of variation and stabilization. *Typological Studies in Language*, pages 55–92, 2008.

J Rodd, G Gaskell, and W Marslen-Wilson. Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, 46(2): 245–266, 2002.

C Rohrdantz, A Hautli, T Mayer, M Butt, DA Keim, and F Plank. Towards tracking semantic change by visual analytics. *Frontiers in Psychology*, 3:304, 2012.

VS Rotenberg. Search Activity Concept: Relationship between Behaviour, Health and Brain Function. *Activitas Nervosa Superior*, 51(1):12–44, 2009.

S Ruiz-Correa, D Santani, B Ramirez-Salazar, and I Ruiz-Correa. Sensecityvity: Mobile crowdsourcing, urban awareness, and collective action in mexico. *IEEE Pervasive Computing*, 16(2):44–53, 2017.

S Rumi, K Deng, and F Salim. Crime event prediction with dynamic features. *EPJ Data Science*, 7, 2018.

S Sachdeva and S McCaffrey. Using social media to predict air pollution during California wildfires. *Proceedings of the 9th International Conference on Social Media and Society*, pages 365–369, 2018.

C Saedi, A Branco, JA Rodrigues, and JR Silva. WordNet Embeddings. *Proceedings of the 3rd Workshop on Representation Learning for NLP*, pages 122–131, 2018.

E Sagi, S Kaufmann, and B Clark. Semantic density analysis: Comparing word meaning across time and phonetic space. *Proceedings of the EACL 2009 Workshop on GEMS: Geometrical Models of Natural Language Semantics*, pages 104–111, 2009.

D Sandra and S Rice. Network analyses of prepositional meaning: Mirroring whose mind-the linguist's or the language user's? *Cognitive Linguistics*, 6:89–130, 1988.

KM Scaplen, AA Gulati, VL Heimer-McGinn, and RD Burwell. Objects and Landmarks: Hippocampal Place Cells Respond Differently to Manupulations of Visual Cues Depending on Size, Perspective, and Experience. *Hippocampus*, 24:1287–1299, 2014.

F Scarselli, M Gori, AC Tsoi, M Hagenbuchner, and G Monfardini. The graph neural network model. *IEEE Transactions on Neural Networks*, 20(1):61–80, 2009.

B Schölkopf and HA Mallot. View-Based Cognitive Mapping and Path Planning. *Adaptive Behavior*, 3(3):311–348, 1995.

R Schwitzgebel. The performance of Dutch and Zulu adults on selected perceptual tasks. *Journal of Social Psychology*, 57:73, 1962.

JR Searle. Intentionality and the use of language. *In: A Margalit (Eds.) Meaning and Use, Reidel Publisher.*, pages 181–197, 1979.

J Silvertown. A new dawn for citizen science. *Trends in Ecology and Evolution*, 24 (9):467–471, 2009.

DW Sims, AM Reynolds, NE Humphries, EJ Southall, VJ Wearmouth, B Metcalfe, and RJ Twitchett. Hierarchical random walks in trace fossils and the origin of optimal search behavior. *Proceedings of the National Academy of Sciences*, 111: 11073–11078, 2014.

T Sit. Using the Weather Research and Forecasting Model to validate and improve high resolution weather forecasts around cities. *WPCCS, University of Warwick*, 2017.

PK Smith and Y Trope. You Focus on the Forest When You're in Charge of the Trees: Power Priming and Abstract Information Processing. *Journal of Personality and Social Psychology*, 90(4):578–596, 2006.

M Sophocleous. Interactions between groundwater and surface water: The state of the science. *Hydrogeology Journal*, 10(1):52–67, 2002.

AW Spirn. *The Language of Landscape.* Yale University Press., 1998.

BJ Stankiewicz and AA Kalia. Acquisition of structural versus object landmark knowledge. *J Exp Psychol Hum Percept Perform*, 33(2):378–390, 2007.

DE Stansbury, T Naselaris, and JL Gallant. Natural scene statistics account for the representation of scene categories in human visual cortex. *Neuron*, 79(5): 1025–1034, 2013.

A Stopczynski, V Sekara, P Sapiezynski, A Cuttone, MM Madsen, and JE Larsen. Measuring large-scale social networks with high resolution. *PLoS ONE*, 9:e95978, 2014.

V Strang. *The Meaning of Water.* Berg Publishers., 2004.

TD Sweeny and D Whitney. Perceiving crowd attention: Ensemble perception of a Crowd's gaze. *Psychological Science*, 25(10):1903–1913, 2014.

L Talmy. Force dynamics in language and cognition. *Cognitive Science*, 12(1): 49–100, 1988.

Z Tang, L Zhang, F Xu, and H Vo. Examining the role of social media in California's drought risk management in 2014. *Natural Hazards*, 79(1):171–193, 2015.

JR Taylor. *Linguistic Categorization: Prototypes in Linguistic Theory.* Oxford: Clarendon Press., 1995.

MH Tewolde and JC Smithers. Flood routing in ungauged catchments using Muskingum methods. *Water*, 32(3):379–388, 2006.

C Thang-Duong, R Lebret, and K Aberer. Multimodal classification for analysing social media. *arXiv:1708.02099*, 2017.

B Thomee, DA Shamma, G Friedland, B Elizalde, K Ni, D Poland, D Borth, and L-J Li. YFCC100M: The New Data in Multimedia Research. *Communications of the ACM*, 59(2):64–73, 2016.

N Tkachenko, S Bricker, and S Jarvis. To dig or not to dig? place and perception in subsurface housing. *Proceedings of the Institution of Civil Engineers - Engineering Sustainability*, 171(4):211–220, 2015.

N Tkachenko, R Procter, and S Jarvis. From Kontrollverlust to Kontrollgewinn: Rediscovering a meso-level normative hierarchy for integrated water management in England using on-line activities. *Tenth International AAAI Conference on Web and Social Media*, 2016a.

N Tkachenko, R Procter, and S Jarvis. Predicting impact of urban flooding using open data. *Royal Society Open Science*, 3, 2016b.

N Tkachenko, R Procter, and S Jarvis. Predicting floods with Flickr tags. *PLoS ONE*, 12(2):e0172870, 2017a.

N Tkachenko, R Procter, and S Jarvis. How do eyewitness social media reports reflect socio-economic effects of natural hazards?. *In: Ciampaglia G., Mashhadi A., Yasseri T. (eds) Social Informatics.* SocInfo 2017. Lecture Notes in Computer Science, vol 10540. Springer, Cham. 2017b.

N Tkachenko, A Zubiaga, and R Procter. WISC at MediaEval 2017: Multimedia Satellite Task, Working Notes. *Proceedings of MultimediaEval 2017*, 2017c.

W Tobler. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46:234–240, 1970.

A Townsend. *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia.* New York: W.W. Norton and Company., 2013.

M Del Tredici, R Fernández, and G Boleda. Short-term meaning shift: An exploratory distributional analysis. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, page 2069–2075, 2019.

B Tversky. *Cognitive maps, cognitive collages, and spatial mental models.* In: Frank A.U., Campari I. (eds) Spatial Information Theory A Theoretical Basis for GIS. COSIT 1993. Lecture Notes in Computer Science, vol 716. Springer, Berlin, Heidelberg., 1993.

MD Vernon. *Perception Through Experience*. London, Methuen., 1971.

CJ Wakslak, PK Smith, and A Han. Using Abstract Language Signals Power. *Journal of Personality and Social Psychology*, 107(1):41–55, 2014.

D Waller and AR Richardson. Correcting distance estimates by interacting with immersive virtual environments: Effects of task and available sensory information. *J Exp Psychol Appl*, 14(1):61–72, 2008.

D Wang, M Zhou, S Ali, P Zhou, Y Liu, and X Wang. A novel complex event processing engine for intelligent data analysis in integrated information systems. *International Journal of Distributed Sensor Networks*, 12(3), 2016.

Q Wang and JE Taylor. Patterns and limitations of urban human mobility resilience under the influence of multiple types of natural disaster. *PLoS ONE*, 11(1): e0147299, 2016.

RF Wang and ES Spelke. Updating egocentric representations in human navigation. *Cognition*, 77(3):215–250, 2000.

Y Wang and K Cao. Context-aware Complex Event Processing for event cloud in Internet of Things. *International Journal on Advances in Information Sciences and Service Sciences*, 5:1–6, 10 2012.

Y Wang and JE Taylor. Coupling sentiment and human mobility in natural disasters: A Twitter-based study of the 2014 South Napa Earthquake. *Natural Hazards*, 92: 907, 2018.

DJ Watts and SH Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

J Weng, Y Yao, E Leonardi, and BS Lee. Event detection in Twitter. *Proceedings of 11th International Conference on Weblogs and Social Media*, pages 1–21, 2011.

R Westerholt, B Resch, FB Mocnik, and D Hoffmeister. A statistical test on the local effects of spatially structured variance. *International Journal of Geographical Information Science*, 32(3):571–600, 2018.

GF White. *Natural Hazards: Local, National, Global.* Oxford University Press., 1974.

A Wierzbicka. *'Prototypes save': On the uses and abuses of the notion of 'prototype' in linguistics and related fields. In SL Tsohatzidis, ed., Meanings and prototypes: Studies in linguistic categorization.* London: Routledge., 1990.

ME Wilkinson, PF Quinn, NJ Barber, and J Jonczyk. A framework for managing runoff and pollution in the rural landscape using a Catchment Systems Engineering approach. *Science of the Total Environment*, 468-469:1245–1254, 2014.

ME Wilkinson, EB Mackay, PF Quinn, M Stutter, KJ Beven, CJA MacLeod, MG Macklin, Y Elkhatib, B Percy, C Vitolo, and PM Haygarth. A cloud based tool for knowledge exchange on local scale flood risk. *Journal of Environmental Management*, 161:38–50, 2015.

P Wohlleben. *The Hidden Life of Trees: What They Feel, How They Communicate - Discoveries from A Secret World.* Greystone Books., 2018.

Z Wu and M Palmer. Verbs semantics and lexical selection. *Proceedings of the 32nd Annual Meeting on Association for Computational Linguistics*, pages 133–138, 1994.

K Xu, W Hu, J Leskovec, and S Jegelka. How Powerful are Graph Neural Networks? *International Conference on Learning Representations*, pages 1–17, 2019.

Y Yang and J Eisenstein. A log-linear model for unsupervised text normalization. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, 2013.

PN Zachary. How small is it? Comparing indices of small worldliness. *Network Science*, 5(1):30–44, 2017.

F Zaidi. *Small world networks and clustered small world networks with random connectivity.* Social Network Analysis and Mining, Springer., 2012.

L Zhao, H Li, Y Sun, R Huang, Q Hu, J Wang, and F Gao. Planning Emergency Shelters for Urban Disaster Resilience: An Integrated Location-Allocation Modeling Approach. *Sustainability*, 9(11):2098, 2017.

B Zhou, A Lapedriza, J Xiao, A Torralba, and A Oliva. Learning Deep Features for Scene Recognition using Places Database. *Advances in Neural Information Processing Systems 27 (NIPS)*, 2014.

FJ Zuuren and HM Wolfs. Styles of information seeking under threat: Personal and situational aspects of monitoring and blunting. *Personality and Individual Differences*, 12:141–149, 1991.