What is Known About Climate Change? A Knowledge Graph Approach

Terry Zhang

Linacre College

University of Oxford

Submitted in partial completion of the

Master of Science in Social Data Science

Trinity 2021

Acknowledgements

I am extremely grateful to my supervisors, Dr. Luciano Floridi and Dr. Scott H. Hale, for their guidance and wisdom. Dr. Floridi provided valuable advice on the conceptualization of my theoretical framework and topic. Dr. Hale was instrumental in assisting with the development and implementation of the methodology. This thesis would not be where it is without their guidance.

I am also extremely grateful for the support of my friends. The completion of my degree and dissertation would not have been possible without their camaraderie and support through the difficult isolating times of lockdown. From lockdown walks to enjoying the sunsets at Port Meadow, I will always be grateful for their friendship and cherish the memories we have shared.

I would also like to thank my family for their unwavering belief in me. It was not easy being so far from them, but my time away was made easier knowing they will always have my back and believes in me. For that, I am deeply grateful.

Abstract

The latest findings from the Intergovernmental Panel on Climate Change Sixth Assessment Report retells a story already told in 1979 in greater detail — that climate change is happening because of human causes. However, the progress on addressing the climate crisis hardly corresponds to the media's nearly 40 years of news coverage. A widely cited barrier to progress is the lack of knowledge to comprehend the various interacting aspects of climate change. As a dynamic complex system, climate change is a deeply relational systems-level problem that requires communication and mental models capable of communicating system complexities. To establish the extent of relations and knowledge formations by past communications, this paper offers a methodological innovation enabling network analysis of unstructured texts. This research developed a reproducible automated semantic network approach using natural language processing and open information extraction tools to construct knowledge graphs (KGs) on climate change news coverage in 5 countries using news articles ($N = 19,684$) from 1990 to 2020. Country comparisons were carried out cross-sectionally and longitudinally. The findings show that the KGs are structurally similar while static but differed substantially in longitudinal patterns. Countries performing better in climate commitments saw a uniform increase in the salience and embedded of concepts, while in those less progressive countries, concepts become salient sporadically. The study concludes by recommending a concerted communication model that gives equal and parallel issue-attention to the causal, action, and physical characteristic and consequence frames of climate change to facilitate understanding of the system complexity of climate change.

Contents

List of Figures

List of Tables

List of Abbreviations

CC . . . . Climate change

GHG . . . . . Greenhosue gas

KG . . . . Knowledge Graph, an implementation of the semantic network emphasizing on ontology and automated construction methods.

IE/Open IE . Information extraction or open information extraction. The extraction of structured data from unstructured input such as text.

NPs . . . . . . Noun phrases

The best time to stop climate change was 40 years ago.

The second best time is now

— Chinese proverb (parody)

1 Introduction

Contents

1.1 Motivation

Everything that we needed to know about climate change, including how to stop it, has been known since 1979 (Rich, 2019). However, the recently published Sixth Assessment Report (AR6), the most authoritative document on the current state of the climate, by the Intergovernmental Panel on Climate Change (IPCC) was called a "code red for humanity" by UN Secretary-General António Guterres (IPCC, 2021). The increasingly common extreme weather events and other consequences of climate change is a difficult reality that must be communicated effectively and understood by public and political actors to mitigate the worst of impacts. Unfortunately, however, climate change and its nearly four decades of media coverage have demonstrated a lacklustre outcome, with climate denialism, skepticism, and cognitive dissonance still prevalent in the public discourse (Dryzek et al., 2011; Farrell, 2019; Ruiu, 2021).

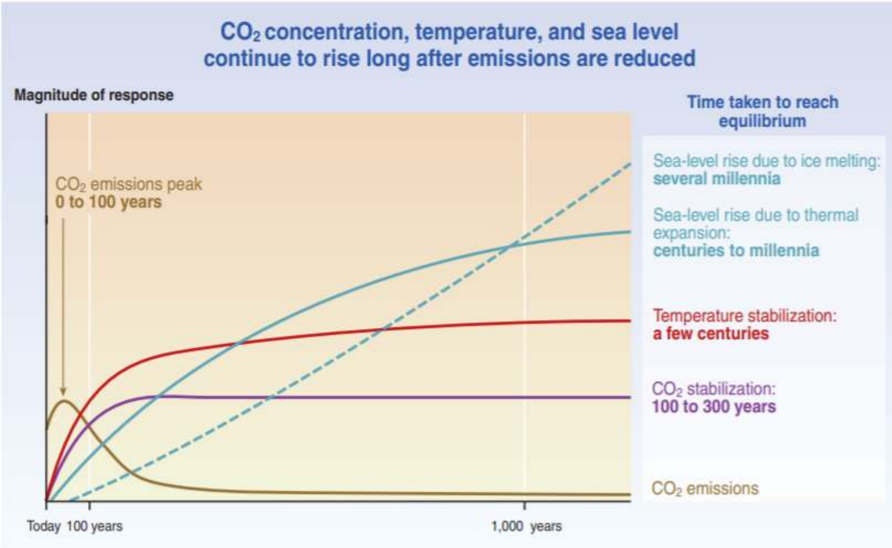


Figure 1.1: The inertia effect of the climate system given a peaked and reduced emissions of anthropogenic sources of CO_2. (IPCC, 2001)

As the world progresses towards a greater climate change, communication cannot become complacent and lessons from the long history of climate change communication must be extracted to inform future strategies. Warning of the effects and threats of climate change from anthropogenic sources of greenhouse gases (GHG) has, unfortunately, already been raised since decades ago by the scientific community (Change et al., 2007; Houghton, 1996; National Research Council, 1979; Sawyer, 1972). However, to date, "only timid measures have been put in place to reduce the use of fossil fuels and CO2 emissions" (Manzanedo & Manning, 2020, p. 2). In the AR6, the current state of the climate continues to worsen, with global surface temperature observed to be already over $1°C$ higher than 1850-1900 pre-industrial levels (IPCC, 2021). This has led to more frequent extreme heat events, increased global precipitation, and likely increased compound extreme events (e.g. concurrent heatwaves and droughts, floods and fires) (IPCC, 2021). To that end, there is an imperative to understand what has been said about climate change.

Climate change has been a difficult concept to comprehend given its large geographical and time-scale consequences and complex nature of interactions (Moser, 2010; Sterman, 2008). The complexity and scale of climate change at a systems level are cited as an ongoing challenge for communicators and a possible explanation of the stagnancy in mitigation and adaptation efforts (Sterman, 2008). The climate is not a single system but rather a *system of systems* consisting of interacting subsystems such as the meteorological (e.g. wind, rain, humidity), ecological (e.g. animals, plants, food chains), and human systems (e.g. economy, society, transportation). Due to the complex interacting nature and magnitude of individual subsystems, climate change effects exhibit high "inertia" in responding to interventions (IPCC, 2001, p. 16). To illustrate, even if CO_2 emissions peak and stabilize within the next 100 years, Earth's surface temperature will take centuries to stabilize, and sea levels will continue to rise for up to millennia (Fig. 1.1). This means that by the time adverse changes have been detected in the climate and ecological systems at an alarming level, it is likely too late to address the contributing system – the human system. Given the importance of communicating the systems nature of climate change, the overarching research question this thesis raises is "*How has media communicated the interconnected nature of climate change over the last four decades?*"

The media remains the trusted source and actively provides up-to-date scientific and risk communications for climate change. Its importance is accentuated by the need to communicate and frame the unobtrusive nature of climate change's systems interaction to facilitate public understanding (Schäfer & O'Neill, 2017). In practice, mass media scientific and risk communication is often less complete and coherent and rarely communicates the exact scientific consensus message (see M. T. Boykoff & Boykoff, 2007). Tracking the coverage of climate change across countries and over time has been done mostly at a single-country level (e.g. Carvalho & Burgess, 2005; J. C. E. Liu & Zhao, 2017; Takahashi & Meisner, 2013). Cross-country studies relied on simple counts of article frequencies or manual coding of themes (e.g. Brossard et al., 2004; Schmidt et al., 2013). More recently, content analysis-based studies have also appeared but continued to rely on conventional coding approaches, which cannot be scaled to understand historical communications across time and countries (e.g. Painter, 2013). The early works have laid the foundation for studying climate change communication by credence to the central concepts, including issue-attention and media framing. However, in this thesis, I posit that to reflect the interconnected nature of climate change and its communication more accurately, there is firstly a need for methodological innovation.

1.2 Contribution

Enabled by natural language processing (NLP) and network science approaches, this paper presents a scalable analysis pipeline to constructing a structured relational representation of historical discourses from large unstructured text data. The pipeline can easily be transferred to other domains. Within this paper, climate change communication is analyzed to describe cross-country patterns in network structures of the KGs longitudinally and cross-sectionally. Significant patterns from past communication are valuable for informing the development of risk communication models.

The paper is organized with the next section providing a literature review of relevant works in climate change communication. Specifically on the concepts of issue-attention, communication and mental models, and framing (Chapter 2). The following section describes the theoretical framework of frame semantics and its implementation using knowledge graphs (Chapter 3). Following is a detailed description of the methodological framework developed within this paper that will provide valuable information enabling greater relational analysis of unstructured texts (Chapter 4). Lastly, the results are presented in Chapter 5 followed by discussion and conclusion in Chapter 6.

Science and technology revolutionize our lives,
but memory, tradition and myth frame our response.

— Arthur M. Schlesinger Jr. (Schlesinger Jr., 1986)

2
Background

Contents

2.1 Introduction

The coverage of climate change in mass media, also called climate change communication, has been studied from many perspectives over the past decades. There is a considerable amount of literature published across disciplines that focuses on climate change communication. The majority of these studies emphasized the media outputs of climate change journalism (Schäfer & Painter, 2021). To answer the research question on structural climate change knowledge differences, four concepts from within climate change communication [1] literature will guide this

[1]climate change journalism and climate change communication are used interchangeably in this paper. Distinctions are made in some literature.

review: the concepts of issue attention, climate change knowledge, communication models, and media framing.

2.2 Issue-Attention

A cyclical pattern of coverages on social issues was proposed as the *issue-attention cycle* by Downs in 1972. Downs suggested that while not all social problems will adhere to such a pattern, those exhibiting three specific characteristics are more likely to. Firstly, the sufferers of the social issue are proportionally smaller than those unaffected. Secondly, the solution requires sustained effort at a systemic level due to the deeply socially embedded nature of the problem. Lastly, the issue is intrinsically unexciting and requires dramatic coverage. Thus, the issue attention cycle is theorized to follow five waxing and waning stages: *1) pre-problem stage, 2) alarmed discovery and euphoric enthusiasm, 3) realization of the cost of solutions, 4) decline of public interest, and 5) the post-problem stage* (Downs, 1972).

Since its proposal, the issue-attention cycle hypothesis has been studied in numerous national and temporal contexts within climate change communication (see Schmidt et al., 2013). The hypothesis was tested empirically, and the issue-attention cycle pattern was found to be present for news coverage of global climate change by McComas and Shanahan (1999) for *The New York Times* and *The Washington Post.* Periods of increasing attention to global climate change were characterized by a more prominent focus on the dangers and consequences. In contrast, the maintenance period of attention was found to feature scientific controversies more prominently (McComas & Shanahan, 1999). Studies since the 2000s have begun looking at issue-attention beyond the US, and the pattern becomes less cyclical but broadly demonstrates some waning and waxing of attention regardless. For example, the issue-attention cycle was found in Peru between 2000 to 2010. Issue attention rose from 2006 until a peak in 2007-2008 followed by a decrease in 2009-2010 (Takahashi & Meisner, 2013), in Germany, France, UK, and US between 2000 to 2010 with an overall upward trend in attention and distinctive peaks that disrupted a smooth cyclical pattern during international policy events (e.g. Kyoto Protol, Conference of the Parties) (Brossard et al., 2004; Grundmann & Scott, 2014). The authors have also offered theoretical explanations to the forces driving the media issue-attention cycle within the issue-attention literature.

2.2.1 Explaining Issue-Attention

An extension of Downs (1972)'s original theory by McComas and Shanahan (1999) sought to describe such forces behind the issue-attention cycle. Downs originally theorized that the innate characteristics (i.e. the three specific characteristics) of an issue drives the cycle. In other words, media coverage increases because climate change is inherently seen as a potentially dramatic issue that requires substantial efforts to overcome. McComas and Shanahan instead explained that "...it is not only the frequency of coverage but also the character and form of that coverage that helps to draw public attention to environmental issues" (1999, p. 53). This explanation is useful because it brought the role of media beyond reporting and in shaping the public discourse and perception of climate change through characterizations.

Media characterization can be understood as the way an issue is presented within the news media by journalists. This may be a presentation of an issue as pressing, controversial, or passing through stylistic characterizations such as personalization, dramatization or minimization to fit the narrative cycle of news stories (M. T. Boykoff & Boykoff, 2007; McComas & Shanahan, 1999). The extent of media characterization in US prestige media was made salient when M. T. Boykoff and Boykoff (2004) found that the journalistic norm of balanced reporting created a controversial representation of climate change despite broad scientific consensus. By adhering to journalistic norms such as objectivity and balance, news media have created a "false balance" that may have lead to a polarized public and political discourse (M. T. Boykoff & Boykoff, 2004, 2007). Additionally, Brossard et al. (2004) suggests that cultural contexts are an additional dimension within media characterization that must be considered when studying climate change communication. Specifically, coverage in France covered environmental policy conflicts between the US and Europe, while US coverage focused on controversies between scientists and politicians

(Brossard et al., 2004). Thus, while media characterization may be considered an internal force within the media landscape that drives the issue-attention cycle, studies have also sought to explore external forces.

Earlier studies of the mechanisms behind issue-attention have made explanatory claims, but few offered empirically-grounded investigations until recently (Schäfer & Painter, 2021). External forces are events and developments within the climate change movement that drives media attention because of their newsworthiness. For example, the Intergovernmental Panel on Climate Change (IPCC) fifth assessment release in 2014 and the Conference of the Parties (COP15) to the United Nations Framework Conventions on Climate Change in 2009 were events that spiked media attention (Moser, 2016; Schmidt et al., 2013). Indeed Schäfer et al. (2014) found that political events such as the IPCC and COP15 are a driver of media attention in Australia, Germany, and India. Moser (2016) also suggests in her review that the increasingly frequent extreme weather events such as Superstorm Sandy and Typhoon Haiyan have driven the issue attention, but Schäfer et al. (2014) found extreme weather events to only be a relevant driver in Germany. Similarly, across 41 countries of varying economic development statuses, Climate Change Risk Index appears to have a negligible effect on the media attention to climate change (Barkemeyer et al., 2017). The recent findings echo the observations of M. T. Boykoff and Boykoff (2007) that climate change attention tends to follow policy developments more than a "natural" cycle. This suggests that, generally, political events are a greater driver of issue-attention than actual climatic events.

Two structural gaps remain within the issue-attention literature that this paper will contribute towards: the majority of the existing literature remains single-country focused, and few papers currently examine the structural difference of issue-attention across countries (Barkemeyer et al., 2017). The scarcity of empirical cross-country literature has been a widely cited gap for issue-attention literature and, broadly, climate change communications (Barkemeyer et al., 2017; M. T. Boykoff & Boykoff, 2004; Schmidt et al., 2013). While single-country studies are valuable for explaining country-specific media characterizations and audience values, comparative cross-country empirical studies may better explain effective climate change communication elements. The long history of studies describing single-country issue attention provides the foundation for more explanatory cross-country studies. In addition, cross-country studies will be crucial to understand structural barriers and national contexts that may empower or hinder a country's media's role in communicating climate change (Anderson, 2009). Thus, properties of the issue-attention cycle are expected within the longitudinal patterns of the KGs.

2.3 Communicating Climate Change Knowledge

Knowledge of climate change is fundamental to achieving timely and effective mitigation and adaptation. Early researches have described the importance of knowledge in behaviour changes and pro-climate political participation (Bord et al., 2000). In 2020, news media remains the most widely used source of information for climate change across 40 developing and developed nations (N. Newman et al., 2020). However, despite the important role of knowledge in the climate change movement, there has been limited investigation into the systematic differences in the presentation of climate change knowledge by the media in different countries (Tobler et al., 2012). As a result, the bulk of the studies within this strand of climate change communication studies have focused on the public understanding and perception of climate change rather than measuring knowledge in a standardized and comparable way.

This review will first focus on climate change knowledge studies approached from a public perception perspective – that is, studies concerning the public understanding of climate change knowledge. Studies on public perception of climate change have a rich body of literature, and there exist both quantitative and qualitative studies (Capstick & Lorraine, 2018). Early studies from the 1980s and 1990s were mostly small-scale surveys but established a baseline for longitudinal comparisons Capstick et al., 2015. In a review by Capstick et al. (2015), climate change knowledge in the early 1980s and 1990s was characterized by misconceptions and lack of knowledge on the anthropogenic causes. For example, a study in the US during that time period found that laypeople conflated the ozone-depleting Chlorofluorocarbons (CFCs) with GHGs and generally having a poor understanding of the anthropogenic causes of climate change (Bostrom et al., 1994). A similar study found that there were also misconceptions and a lack of knowledge in the causes of climate change in the appropriate responses in Sweden (Löfstedt, 1991). These misconceptions, such as between CFCs and GHGs, have even led to beliefs in reducing aerosol spray cans usage instead of energy conservation measures (Bostrom et al., 1994; Löfstedt, 1991). Although somewhat reassuringly, most respondents from the 1994 study correctly mentioned some aspects of the consequences of global warming (Bostrom et al., 1994).

Even in the early studies, Bostrom et al. noted that "despite widespread media coverage of global climate change and related issues, lay mental models of global climate change suffer from several basic misconceptions" (1994, p.968). Indeed, in a replication of Bostrom et al. (1994)'s study 17 years later, Reynolds et al. found that while some understanding have improved, "[the] laypeople's mental models now have changed surprisingly little since 1992" (2010, p.1534). The persistence of misconceptions on climate change despite years of increased media attention and growing scientific consensus may be rooted in several causes.

2.3.1 Communication Models

A much debated question is whether knowledge and facts are important in scientific and risk communication models to elicit public participation in climate change actions (see Bord et al., 2000; cf. Kahan et al., 2012). In scientific communication, the dissemination of scientific knowledge follows a longstanding "information deficit" model, which assumes the audience lacks the scientific information and proper understanding to address an issue (Wynne, 2006). The efficacy of this model has since been debated for the exact reason found in Reynolds et al. (2010)'s study and the growing scientific skepticism amongst the public (Wynne, 2006). In fact, in a survey of US adults, Kahan et al. (2012) found that individuals with higher scientific literacy perceived climate change as a less serious issue than those with lower scientific literacy, thus rejecting the "information deficit" hypothesis; instead, the authors found cultural worldviews to be an important predictor for climate change perception. If the findings of Kahan et al. are accurate, then the importance of cultural and personal values outweighs that of knowledge in raising the public perception of knowledge. Wynne (2006) and several other studies (see Pearce et al., 2015a, p.619) have proposed a more dynamic and interactive approach to climate change communication. Specifically, a deliberative approach that emphasizes emotional engagement and is contextualized on local issues and cultural values.

More recently, studies have refuted the contextualized value-driven communication model and again accentuated the importance of knowledge dissemination in climate change communication. In a detailed investigation of the importance of knowledge and value for explaining climate change perception across six developed nations, Shi et al. found that knowledge remains "...an important predictor of cross-national public concern even when we control for different value orientations" (2016, p.761). Shi et al.'s conceptualization of climate change knowledge as three distinctive domains, physical characteristics, causes, and consequences, provided more granular insight into understanding the role of knowledge. It appeared that knowledge about climate change causes was significant for public perception, while physical knowledge (e.g. effects of CO_2 versus methane) were even negatively correlated to concern about climate change. This result is similar to the earlier findings (Tobler et al., 2012), that knowledge about climate change causes was most important to risk perceptions. Therefore, not all domains of climate change knowledge contribute evenly to public perceptions and bring into question the role of media coverage in shaping public knowledge.

2.4 Media Framing of Climate Change

While knowledge and communication models offer some theoretical explanation to the persistence of climate change misconceptions, the way media covers the issue is another variable that must be considered. Throughout early climate change communication studies, scholars have noted the role of the media in translating and disseminating scientific knowledge to the public and political sphere (Bostrom et al., 1994; M. T. Boykoff & Boykoff, 2004). The interdisciplinary nature of climate change knowledge means many knowledge domains and a correspondingly large pool of information to draw stories from. As Shi et al. conceptualized three domains mentioned previously, there is additional knowledge that the media may include or exclude when covering an issue. The intentionally selective representation of information by the media has been studied as media "frames "(Moser, 2010). Framing is a core concept to understand how the media can shape the knowledge of the public.

There are several conceptualizations of frames within framing scholarship. Matthes (2007) distinguished two general frames: *formal-stylistic* and *content-oriented*. Formal-stylistic frames describe the structural and formal representation of information, whereas content-oriented frames focus on the content. The most widely studied perspective within climate change communication is content-oriented frames (Schäfer & O'Neill, 2017). Additionally, content-oriented framing is the most relevant frame for this study because of its presumed role in shaping the information and knowledge audiences receive.

The media is rarely a simple, neutral communicator. Instead, strategically selective coverage decisions, or framing, are commonly employed to achieve some predetermined goals within the expected audience (Moser, 2010). These goals may be readership and revenue in for-profit media organizations like those in the USA, a national agenda for state media, or value-driven goals such as environmental protection. Entman describes framing, specifically content-oriented frames, as "...select[ing] some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation" (1993, p.52). More concretely, framing is the deliberate inclusion or exclusion of some information or knowledge about an issue in the media.

There is a great amount of variability in how climate change knowledge is framed across countries (Moser, 2016). For example, in Sweden Olausson (2009) found that collective responsibility and certainty frames were dominant, and the media was reluctant to adopt a scientific uncertainty frame. Similar frames were found to be present in France (Brossard et al., 2004). Conversely, the scientific uncertainty framing of climate change was much more dominant in the USA (Antilla, 2005). Climate change is often framed through natural impacts in natural disaster-prone countries such as The Philippines, India, and the Gambia (Vu et al., 2019). Considered from a knowledge perspective, it can be presumed that a national media frame that provides more attention to scientific uncertainty contributes to the persistence of climate misconceptions or structural knowledge gaps within the country's public. The cross-national differences in framing are expected to create structural differences in the knowledge available to the audiences.

Through experimental researches, frames have been studied to understand their effects on audience perceptions of climate change. For example, in an experiment conducted in UK and Italy, Spence and Pidgeon (2010) found that by emphasizing the positive outcomes of climate measures (gain frames), there were improved information recall and increased the perceived severity of climate change than those presented with the negative consequences of inaction (loss frames). SIMILAR RESEARCH The experimental evidence shows that the same knowledge of climate change action presented in different frames, some selective presentation of information, created disparate outcomes in perception. Therefore, structural differences in public perception due to different knowledge compositions across countries will likely be observed.[2].

The task of comparing frames cross-nationally is faced with methodological difficulties. Frames have been widely studied from a range of methodological approaches, but it "...is a *fundamentally constructivist concept*" (Schäfer & O'Neill, 2017, p.31). Using a set of master (generic) frames, such as the widely-adopted five generic frames by Semetko and Valkenburg (2000), comparability can be achieved across countries. However, comparable results are scarce among the

[2]This expectation is stated under the assumption that national perceptions are isolated units for the purpose of this paper. In reality, international news flow is a considerable force (see Takahashi & Meisner, 2013)

majority of topical frame research because of differences in coding decisions and the definition of frames. More recently, computational linguistic approaches have gained popularity due to their ability to work with big corpora in a replicable manner. The most commonly adopted approaches are topic models such as latent Dirichlet allocation (LDA) or word clusters (Schäfer & O'Neill, 2017). However, computational linguistic techniques, while powerful, are questionable operationalization of the framing concept due to their unsupervised nature. Thus, they are rarely presented as frame studies.

The field of frame analysis can undergo considerable methodological improvements to investigate the role of framing in shaping the public's climate change knowledge. However, current research relies on manual constructivist approaches that reduce replicability and produce few cross-nationally comparable results. The methodological limits of manual content analysis approaches may have also prevented widespread investigation into the change in media frames over time (for an exception see Vu et al., 2019). The longitudinal understanding of media frames and knowledge presentations combined with climate change perception data can provide an enhanced picture of the role knowledge has in shaping the public and political agenda. To that end, this paper will seek to address the gap in current content frame studies by employing a replicable method that can provide longitudinal and cross-nationally comparable results.

Consider all things connected.

3
The Frameworks

Contents

This section will introduce the theoretical concept of frames and describe the analytical framework using semantic networks implemented as knowledge graphs.

3.1 Theoretical Framework

3.1.1 Framing and Understanding

Research focused on understanding framing has grown ever more important as a growing body of literature has shown framing effects on instigating substantial institutional change (Werner & Cornelissen, 2014). As climate change is a deeply rooted institutional issue that, despite years of issue-attention, has resisted effective change leading to the current state of complacency is an area where many questions linger (Rich, 2019). The effectiveness of past frames has clearly been inadequate, and new framing directions must be taken. Reframing by institutional actors to achieve practical impacts and reshape the audience's cognitive frame is documented in literature such as BP's reframing of their role in the 2010 oil spill as causer to solver (Schultz et al., 2012) and the media's framing of Islam following 9/11 leading to its association with terrorism (Ruigrok & van Atteveldt, 2007). As Beckert (2010) stated that "institutions are defined as inter-subjectively shared meanings and thereby become almost indistinguishable from cognitive frames", the latent meanings that past climate change frames have constructed illustrates the path leading to today's cognitive frames and institutional representation of climate change.

3.1.2 Relational Understanding through Frame Semantics

Kovecses (2006) argued that the possession of facts is only part of knowledge and a greater proportion of understanding comes from the "structured mental representation" of underlying concepts, also known as the frame. Akin to a physical picture frame, it is a mental model with which one understands a concept. Such structured mental representation is a product of words' associative nature, as posited by the linguistic theory of frame semantics (Fillmore, 2006). Specifically, frame semantics argues that the semantic (meaning) of concepts and words can only be understood in relation to larger contextual frames such as culture, history, and personal experiences (Fillmore, 2006). The opposite is true that when one understands a word or concept, a specific frame is triggered. To illustrate this, the concept of a "New Year's day" can be understood as the first day of the calendar year. However, a year and a day are only understood in relation to the concept of a calendar year. Therefore, there must also be a year frame within which calendars exist to track the progress of its elements of months and days. Furthermore, New Year's Day depending on the cultural context, this day can be January 1st for Gregorian calendar users or September 11th for Ethiopian calendar users. Thus, there is hardly a universal New Year's Day, and one can only understand it against a calendar frame within a year frame that can exist in two different cultural frames.

Applying this principle to climate change means that individuals understand climate change through different frames and trigger different frames for causes and solutions.

In addition to the relational semantic aspect of frames, Fillmore also posited that through the use of different representations (i.e. words or phrasing), a communicator can evoke a specific frame (mental representation) and provide different interpretations of an objective situation or concept. For example, representing climate change as a product of consumerism may trigger the responsibility frame, whereas calling it an issue of survival may trigger a disaster frame (Olausson, 2009). An assumption that underlies this argument is the availability of the frame that the alternative representation intends to evoke. For example, an individual without knowledge of climate change risks cannot evoke the risk frame without understanding the concepts that underlie the risk frame (e.g. extreme weather and natural disasters). This further reinforces the cross-context differences that exist due to frame availability. One can only evoke frames made available to them by their surrounding contexts, such as culture and media.

An important property of knowledge is that the evoked frame is not necessarily singular, meaning a concept can evoke multiple frames that one knows. For an individual, perhaps an Ethiopian-American, September 11th as a date can be understood through two drastically different frames, with one being a holiday and another terrorism frame. This property of framing is especially important for climate change because of the shortcomings imposed by the mental models and frames through which the complex dynamic systems quality of climate change needs to be understood. Specifically, "we don't [*sic*] understand the process of accumulation (stocks and flows), feedback, time delays, non-linearity and other concepts necessary to understand the dynamics of complex systems such as the climate and economy" (Sterman, 2011, p.816). Informed by the systems complexity nature, there are likely multiple simultaneous mental frames returned by climate change as a concept but with the most salient frames resulting from respective national contexts.

3.2 Research Objective

The following section will discuss the analytical approach to operationalize the concept of frame semantics and relations. It can be understood as a network analysis approach implemented using a semantic network. Recent examples of semantic network analysis can be seen in several domains such as by Kang et al. (2017) to identify central concepts in vaccine opinion groups and a public relations study on BP's crisis communication by Schultz et al. (2012). Motivated by the lack of the application of a network approach to studying the structure and relations of climate change concepts, I raise an initial exploratory research question:

> **RQ1:** What are distinguishable and similar characteristics between the climate change knowledge graphs of the analyzed countries?

Informed by the theory of frame semantics, which proposed that the creation of relations between relevant frames is fundamentally the route of understanding. Under the pretext that understanding is necessary for effective climate actions as shown through the literature review, I extend the argument by proposing that the progress on climate change in a country is reflected by the frames made most salient by the media, which leads to a second research question:

> **RQ2:** What are the most salient frames within each country as measured by relational properties? How do these change over time?

In the context of the relational nature of understanding, perceptions, actions, and attitudes are, therefore, a result of activating the frames constructed within one's environment. A relational approach to understanding framing is thus a more theoretically justified approach than simple counts. Such that leads to the first hypothesis:

> **H2a**: Countries with more progress on climate change have more salient frames relevant to solutions and current policies than causes and consequences.

Furthermore, there is a longitudinal perspective in frame salience as media outlets do not remain static. The earlier a frame stabilizes, the more highly embedded those in the mental model presented by the media. Although the study of actual mental models of individuals is out of scope for this paper, observational findings can provide an initial understanding of frame salience over time. Thereby complementing H1a with a longitudinal perspective, the second hypothesis states that:

> **H2b**: The most salient frames in countries most progressive on climate change see causal, consequential, and action frames stabilized early.

3.3 Analytical Framework

3.3.1 Semantic Networks

The semantic network is a networked representation of knowledge that reflects the relational nature of understanding proposed by the frame semantics theory. The relational representation of concepts and knowledge in semantic networks developed through a rich line of literature into a well-established concept and approach in content analysis (see Doerfel, 1998). In its most basic form, a semantic network follows the conventional node and edge structure. A node is commonly a concept or word that represents an underlying entity. An edge is a specific relation that exists between two nodes extracted via the network construction process. For content analysis applications, semantic networks reveal relational patterns and hidden relational semantics that cannot be identified through conventional word-count or co-occurrence approaches (Carley & Kaufer, 1993). However, despite its early conceptualization, the semantic network did not see widespread applications until recently.

A partial barrier to the wider adoption of semantic networks is the difficulty in constructing them. The process of semantic network construction is considered a task of information extraction (IE), that is, the extraction of structured data from unstructured texts (Maynard et al., 2017). Manual construction by coding and extracting nodes and relations has been a common practice since the early days and is still applied today (see. Doerfel, 1998; Kang et al., 2017; Schultz et al., 2012). Manual approaches yield higher precision but lack the scalability needed to carry out large-scale analysis and miss system-level patterns.

With the recent development of NLP tools and unsupervised machine learning algorithms, a new paradigm of automated IE and network construction have to lead a resurgence in the feasibility of applying semantic networks on content analysis questions. The resurgence has introduced the implementation of semantic networks using knowledge graph techniques. A knowledge graph is a data structure that enables semantic networks at scale. Specifically, knowledge graph-based semantic networks emphasize automatic construction under a formalized ontology. In this paper, I present the present approach to semantic network as a knowledge graph to distinguish the automated construction process (see section 4). To that end, this paper leverages IE and NLP tools to conduct a knowledge graph analysis of climate change communication in five countries under the concept of frame semantics and semantic networks.

3.3.2 Knowledge Graphs

Large strides have been made in knowledge representation formalism towards an effective structured representation of knowledge. One of the most promising approaches within knowledge representation is knowledge graph (KG) Hogan et al., 2021. The conceptualization of KG is rooted in need for a machine-readable representation of "communicable knowledge" in computer-administered instruction systems in 1969 (Kopstein, 1969). Kopstein identified graphs as the ideal representation for structuring communicable knowledge due to its mathematical rigour grounded in set theory while maintaining high human interpretability. However, the term "knowledge graph" was not coined until 1973 by Schneider (1973). The recent rise in popularity for KGs can be attributed to technology firms' growing interest and implementation.

The modern concept of KG gained popularity with the launch of the Google Knowledge Graph in 2012 (Singhal, 2012). KGs also saw implementations in other

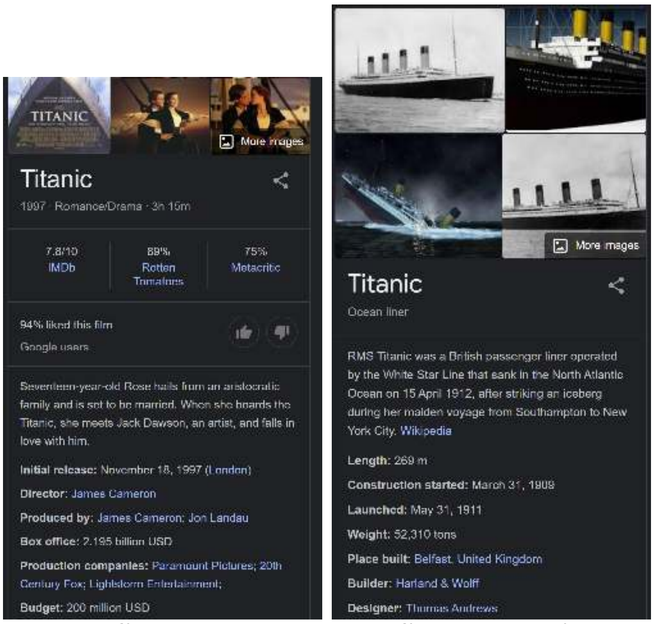


Figure 3.1: Different entities have different sets of relationships.

technology companies such as Amazon's Product Graph[1] and Airbnb's KG[2]. A common benefit named within industry applications is the introduction of semantics into their data structure. The software application understands the entities being asked to retrieve from some metadata and the interrelations between them. For example, Singhal (2012) wrote with the Google KG, execution of searches will no longer be simple keyword matches but also powered by an understanding of entities in the query. So the query "Titanic" will return different results depending on if the user selects the "Ocean Liner" entity type or "Film" entity type, as each entity type has a distinctive set of relationships (fig. 3.1).

In sum, KGs have allowed these commercial applications to store and analyze information beyond simple words but as machine-interpretable facts about the world. This paper is primarily interested in using KGs to reveal the structure of relations and concepts that underlies a country's communication through network analysis.

[1]https://www.aboutamazon.com/news/innovation-at-amazon/making-search-easier

[2]https://medium.com/airbnb-engineering/scaling-knowledge-access-and-retrieval-at-airbnb-665b6ba21e95

3.3.3 Definition

KG is now a well-applied concept in private products and services, yet its formal definition varies greatly within publications Ehrlinger and Wöß, 2016. The difficulty in offering a formal definition results from several closely related but often conflated concepts and the varied implementations of KGs in practice. The purpose of this section is to briefly review the varied definitions and provide a running definition for this paper without a deep analysis of the debate.

The concept of KG was conceived from a line of technologies within knowledge representation and AI. Specifically, KG is closely related to the early and still popular knowledge base (KB) technologies and the Semantic Web (Ehrlinger & Wöß, 2016).

Central to KGs is the relational property between entities. The entities and relationships can be represented in a directed labelled graph $G = \{V, A\}$. V is a set of nodes (vertices) comprising concepts or words representing real-world entities. A is a set of directed edges that are labelled with the relations that two nodes hold. Throughout the rest of the paper, I use the term entities and concepts interchangeably in different contexts, but both refer to the nodes of a KG. Edges are also referred to as relations.

With nodes as the entities and labelled edges as relationships between the entities. A set of triples represents an observation within a KG. For example, the communicable knowledge "greenhouse gases cause climate change" has the relationship "cause" between the entities "greenhouse gases" and "climate change." The example can be represented in a node and edge structure.

Connect all.

4
Methodology

Contents

4.1 Introduction

The creation of knowledge graphs is known as *knowledge graph construction* (KGC), and it is the overarching methodology applied in this thesis. KGC, rather than a single methodology, describes and ensemble approach and variations in the chosen set of techniques are task and resource-dependent (Hogan et al., 2021). The methodology ensemble was implemented in a pipeline structure, with each technique as a component along the pipeline that performs a transformation or analysis of the input data. The KGC pipeline can be modified and improved iteratively to include additional data sources and techniques, enriching the precision and recall of the graph. (Hogan et al., 2021).

KGC from text corpora consisting of unstructured documents such as newspaper articles is a multistage process with the final output of structured relational data extracted from unstructured texts. The variability in human language presents many challenges that were addressed with techniques from NLP and information extraction (IE) (Maynard et al., 2017). First, preprocessing denoised the data and improved the performance of the relation extraction (RE) stage. After successfully extracting relational data, entities and entity clusters were created based on semantic similarity as an alternative approach the labour-intensive named entities recognition (NER) technique found in conventional GKC processes. Using extracted relations and entity clusters, a KG was constructed for each country and time. Lastly, network statistics were calculated, enabling empirical cross-country and longitudinal comparison of entity clusters across graphs. The number of KGC techniques employed in this paper demonstrates the minimal components necessary to carry out scalable cross-country frame analysis within the climate change communication domain.

The ensemble approach of NLP, IE and network statistics techniques has previously been applied by Tangherlini et al. (2020) on conspiracy theories. The authors used KGC components, albeit they did not call it, to develop an automated pipeline for discovering conspiracy theory narrative frameworks within newspaper articles and social media posts. The automated pipeline by Tangherlini et al. represents one of the first applications of a fully automated KGC process in literature. However, while the pipeline within that work produced high precision and recall results, its technical complexity limits the transfer of its application onto additional research questions and for research with fewer resources.

The pipeline described in this section differs from conventional approaches to KGC. Conventional KGC approaches extract only a pre-defined set of relations defined by an ontology. However, given the lack of domain-specific ontology for climate change and the expense of formal ontology construction, I took an *open* IE approach to KGC (Wu & Weld, 2010). An ontology is a generalized description of entity types and the relations its capable of having. To illustrate, the "human" entity type in Wikidata's[1] ontology is described to have the relational attribute of "physically interact" with "natural environment," which can be represented in semantic triple form[2]:

$$(e_1, rel, e_2) \tag{4.1}$$

where e_i are entities and rel are relations as

$$(human, physically\_interact, natural\_environment) \tag{4.2}$$

Using tuple 4.2as the only entity-relation defined in an ontology, from the text "Tom polluted the ocean, and the pollution killed fishes," the triple $(Tom, pollute, ocean)$ would be extracted. However, the triple $(pollution, kill, fish)$ would not be extracted because the entities "pollution" and "fish" were not defined.

The construction of a comprehensive and robust ontology is a labour-intensive process that requires manual inputs. While semi- and fully automatic approaches have been discussed in research (see Maynard et al., 2017), their practical implementations and performance still face considerable challenges (Al-Aswadi et al., 2020). *Open* IE overcomes ontological constraints by extracting all relational instances within a text using unsupervised ML approaches. Furthermore, the ontology-agnostic approach renders the KGC pipeline more accessible to researchers outside of the IE domain.

The KGC pipeline I present further contrasts conventional KGC pipelines to optimize for reproducibility and transfer application onto other social science research questions and domains. Thus, this section can be seen as a practical introduction to constructing KGs for the social sciences. The KGC pipeline employed within this paper does not fully replicate the conventional KGC elements as described by Maynard et al. (2017) and Hogan et al. (2021). However, its reduced components did not prohibit the feasibility of the methodology from answering the research questions and demonstrated the robustness of the KG pipeline for applied research.

[1]https://www.wikidata.org/wiki/Q5
[2]https://www.w3.org/TR/PR-rdf-syntax/

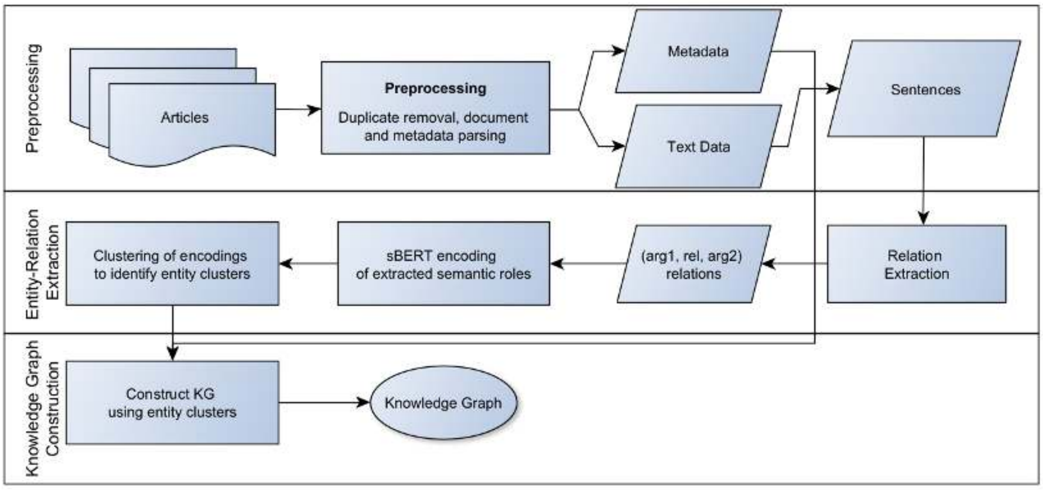


Figure 4.1: An overview of the KGC pipeline employed to construct the climate change knowledge graph.

Furthermore, the current KGC pipeline presented is computationally frugal and was executed using Google Colab Pro with a single GPU and on a single machine with a 6-core Intel i7 CPU at 2.60Ghz with 32GB of RAM.

The following subsections will describe each component within the pipeline presented in the overview presented in the figure 4.1. The total pipeline comprises three stages *preprocessing, entity-relationship extraction* and *knowledge graph construction*, with each stage containing analytical components (i.e. NLP or IE processes).

4.2 Data

The final KGs were constructed using news articles on climate change from five countries between 1990 and 2020 using the Lexis-Nexis database. To represent the global scale of climate change, I selected countries from four continents. The sampled countries were: Australia, Canada, India, UK, USA.

Beyond geographical variability, the selected countries represent a range of national climate change commitments and performances. Using the ranking from the Climate Change Performance Index (CCPI) developed by the non-governmental organization Germanwatch[3], the sampled countries vary drastically in the levels of and progress towards climate commitments. The CCPI tracks country climate

[3]https://ccpi.org

performance by calculating an annual score using a comprehensive set of indicators encompassing GHG emissions, renewable energy, energy use and climate policy (Burck et al., 2021). The latest CCPI scores rank the UK in 5th, India 10th, Australia 54th, Canada 58th, and United States in last at 61st place. The differences in progress may lead to or be a characteristic of structural differences in media content that the KG reconstruction of climate coverage can potentially capture. In Schmidt et al. (2013), national differences in Kyoto Protocol commitments and carbon dependency were similarly used as factors in case selection.

The chosen news outlets were considered on several factors, including historical circulations, to reconstruct national coverage from a longitudinal perspective. At the same time, online media continues to grow amidst the dwindling readership of print media (N. Newman et al., 2020), many traditionally print media such as the New York Times (NYT) have evolved to having a considerable amount of online presence. Fortunately, through the use of a database like Lexis-Nexis, a publication's transition from print to online were captured thus allowing the present method to capture contents that reflect the shifting consumption modes. The inclusion of online media also overcomes the common criticism of climate change research being overly focused on legacy print media (Schäfer & Painter, 2021).

Within each country, two news outlets were chosen for analysis. The choice of two national media attempted to normalize the effects of a single publication's ideological cultures and political alignment (Carvalho, 2007). However, this criterion was limited by the coverages of sources on Lexis-Nexis. For example, Lexis-Nexis coverage of the NYT significantly outnumbered USA Today. The limited rate of collection from Lexis-Nexis was also another constraint on the size of sample data. Lexis-Nexis does not offer volume download options. Thus mass data extraction was carried out using an automated script[4]. Notwithstanding these limitations, the included sources using the proposed KG approach still provided important evidence on structural differences within each country's media coverage.

[4]Data scraping is permitted for non-commercial researches in the UK. See https://www.gov.uk/guidance/exceptions-to-copyright#overview

Table 4.1: Overview of the sampled input data representing climate change coverage in leading national Media outlets.

Countries	Sources	N	Avg. Article Length	Date Range
Australia	Herald Sun	2388	578	1990-02-06 to 2020-12-17
Canada	National Post The Globe and Mail	7122	853	1990-02-06 to 2020-12-30
India	The Times of India (TOI)	1964	505	1990-02-06 to 2020-12-29
United Kingdom	The Times (London)	2924	697	1990-02-06 to 2020-12-28
United States	The New York Times (NYT) USA Today	5286	1209	1990-02-06 to 2020-12-30

The articles were queried using a complex Boolean search string developed by Schmidt et al. (2013) for their study on the media attention of climate change issues in 27 countries.

> (climat* W/5 (chang* OR catastroph* OR disaster* OR transform* OR adjust* OR trend* OR world* OR earth* OR warm* OR heat* OR cool* OR variab*)) OR ((greenhouse* W/3 effect*) OR ((global* OR earth* OR world* OR international* OR hemisphere*) W/5 (warm* OR heat* OR cool* OR chill*)) OR ((temperature*W/5 (global* OR earth* OR world* OR international* OR hemisphere*) W/8 (increas* OR rising* OR rise* OR decreas*))

After refining the queried articles to within the research date range of 1990 and 2020, the final sample contained 19,684 articles. Table 4.1 contains an overview of the sampled data.

4.3 Pre-processing

Pre-processing is a common step within NLP tasks to remove unnecessary noise from documents and corpus. Pre-processing was carried out at both the corpus and document level (first row of fig.4.1).

Duplicate Removal: Duplicate documents at the corpus level is a major source of noise that can cause certain entities and relationships to receive more weight than in reality. Deduplication was carried out using document-level Levenshtein Distance implemented in the Python package `fuzzywuzzy` ().

Metadata Parsing: Metadata such as time and publication were parsed from the Lexis-Nexis article downloads to establish country and time variables for entity-relations.

Sentence Splitting: Sentence splitting of the corpus to attain inputs for the entity-relation extraction stage. Previous NLP works in entity-relations extraction and existing approaches commonly use a sentence-level approach (see Wu & Weld, 2010; also Baker et al., 1998; Palmer et al., 2005).

Multiword Grouping: Multiword grouping was applied to further noise within the data. Combining tokens such as "greenhouse gas" as a single token "greenhouse_gas" has been shown to improve embedding performances over simple tokenization within a domain-specific dataset (Camacho-Collados & Pilehvar, 2018).

Sentence Sampling: This thesis specifically focuses on the relations in regards to climate change and global warming, only sentences containing relevant terms were sampled using a Boolean string for relation extraction.

The output of the preprocessing stage was corpus sentences with their associated article ID to correspond with article metadata.

4.4 Relation Extraction

Relation extraction (RE) describes the function of extracting structured relational data from unstructured texts. It can be considered the most crucial stage in the KGC pipeline because its output is the foundational data structure of KGs. Recalling that a relation is defined as a semantic triple (e_1, rel, e_2) where e_i are entities within the corpus and rel is the entity-relation. The triple is a mathematical representation of the predicate-argument structure found in sentences (Matthews, 2007). Take the sentence "greenhouse gases warmed the Earth" as an example. The subject "greenhouse gases" and object "Earth" are the arguments, and the predicate is the verb "warmed." Arguments are commonly noun phrases (NPs), and predicates are verbs or verb phrases (Matthews, 2007). While higher-order n-ary relations exist and can be extracted (Bach & Badaskar, 2007), I only considered binary relations and extracted relations with two arguments and a single predicate.

Table 4.2: Semantic roles of interest, generalized corresponding grammar roles and explanations.

Semantic Roles	Grammar Roles	Explanation
Arg0	subject	Proto-Agent, one that performs an action
Arg1	object	Proto-Patient, one that is acted upon by Arg0
Arg2	noun phrase	Additional information on action/indirect objcet
V	verb	The predicate or action performed

4.4.1 Semantic Role Labelling

Entities and relations were extracted by recovering the predicate-argument structure from sentences using Semantic Role Labelling (SRL) (He et al., 2017). Semantic roles describe the roles NPs (i.e. arguments) hold regarding the verb (i.e. predicate) within a predicate-argument structure (Bornkessel et al., 2009). The assignment of semantic roles for NPs is dependent on the semantics (meaning) rather than the grammatical position (e.g. subject, object) within a phrase, which allows for an abstract representation of the semantics that underlies syntactic representations. Each verb holds its own set of semantic roles (see Palmer et al., 2005), which, when labelled, allows for arguments to be linked, creating a semantic triple (Bornkessel et al., 2009). *Argument-linking* of semantic roles, thus, can be understood as the underlying mechanism of RE.

The specific semantic role labels used in this research were developed by the Propositional Bank (PropBank) (Palmer et al., 2005). Table 4.2 and 4.3 contain the semantic roles and their patterns extracted within this paper, respectively. *Arg0* and *Arg1* are generalized semantic roles that correspond to the Prototypical Agent and Prototypical Patient roles proposed by Dowty (1991). In addition to the Agent-Patient roles, the inclusion of *Arg2* is to capture more contextual relations for a predicate as seen in Tangherlini et al. (2020).

RE using semantic roles are superior in two major respects to conventional syntactic rule-based approaches. First, it is simpler to implement in analysis and offers superior semantic disambiguation. Early RE techniques relied on syntactic

Patterns	Example Phrases	Example Relations
(Arg0,V,Arg1)	1) Most Republicans regard climate change as a hoax	1) (most republicans, regard, climate_change) 2) (climate_change, cause, levels of hunger)
(Arg0,V,Arg2)	2) Climate change causes alarming levels of hunger	1) (most republicans, regard, climate_change) 2) N/A
(Arg1,V,Arg2)		1) (climate_change, regard, as a hoax) 2) N/A

Table 4.3: Patterns extracted during the RE stage and illustrative examples from the dataset.

features, such as parse tree and parts of speech (POS) tagging, to extract predicate-argument relations (Punyakanok et al., 2008). Rule-based approaches using syntactic approaches may yield higher recall at the expense of manually defining rules and adding labour components to the pipeline. Recent developments leveraging deep learning approaches, specifically bidirectional LSTMs (BiLSTM), enabled end-to-end SRL tasks without inputting syntactic features (He et al., 2017). As the present pipeline took a semantic clustering approach (discussed in the following section) to discover entity clusters, the less precise SRL results by the BiLSTM developed by He et al. (2017) was considered acceptable given its superior accessibility.

Furthermore, RE using semantic roles have the benefit of preserving the semantics of a sentence (Van Valin, 2004). Using the earlier example of "greenhouse gases warmed Earth" and its passive form, "Earth was warmed by greenhouse gases." Syntactic feature-based extractor would detect "greenhouse gases" as the subject in the first sentence but as the object in the passive form. Using conventional *(subject, verb, object)* (SVO) pattern, two different relations and a conflicting direction of interaction would be found from the examples. Conversely, with SRL, "greenhouse gases" holds the same semantic role in both sentences and would be considered the same argument, thus establishing two sets of cohesive relations where the greenhouse gas affected Earth and not vice versa as syntactic approaches would extract.

4.5 Entity Identification

Entities are the nodes within a KG and are another inseparable component of a KG. Entity identification is most commonly carried out using Named Entity Recognition

(NER), an ML approach to identify named entities (NE) within a text (Maynard et al., 2017). A NE is the textual representation or mention of a real-life entity type, such as "Obama" is a NE of the entity person. NER is commonly carried out as a classification task requiring labelled training data, which was not readily available for the climate change domain. As mentioned in section 4.1, this pipeline was developed under the *Open* IE paradigm, which led us to overcome this challenge by innovating and developing a semantic approach to entity identification.

The absence of NER before the RE stage meant argument boundaries were less clearly defined than they would be with labelled NEs and a formal ontology. This led to the extraction of NPs of varying lengths that often contained noises from surrounding words. For example, the words "level of" within example 2 in table 4.3 is noise on top of the actual entity of interest, "hunger." This posed a challenge to readily using the extracted arguments as entities. To overcome this, I leveraged the semantic consistency of the extracted semantic roles. By assuming that the extracted arguments, despite the presence of noise, are meaningful noun phrases that allow unsupervised clustering of arguments into semantically coherent entity clusters $E_i = \{e_1, e_2..., e_i\}$ where every $e_i \in E_i$ is semantically similar. Using the language of Dowty (1991) for prototypical semantic roles within a phrase, I call a cluster of semantically similar noun phrases a *prototypical entity* or *proto-entity*. Semantic clustering for entity discovery have been previously applied by Samory and Mitra (2018) and Tangherlini et al. (2020).

4.5.1 Entity Clustering

Extracted arguments were embedded to a real vector of dimension n using sentence BERT embedding where semantic similarities are preserved (Reimers & Gurevych, 2019). For every e_i it is embedded as a vector v_i and two arguments are considered semantically similar if $||v_i - v_j|| \approx 0$. After embedding m unique arguments as a matrix of (m, n) was outputted, which in this case $(311032, 768)$. While the matrix can be clustered in principle, its large dimensions posed a computational bottleneck. Firstly, conventional clustering algorithms such as k-means require a defined number of clusters which requires an exhaustive search using the Silhouette or Elbow method. With scale of this paper's sample size, this was an unfeasible approach. The application of hierarchical clustering approaches is feasible; however, its implementations become prohibitively slow against high-dimensional data (Y.-c. Liu et al., 2011). Density-based methods such as DBSCAN were also unfeasible because of the explosion in memory that occurs when computing the distance matrix. Calculating the cosine distance of a matrix with $311,032$ rows of 32-bit integers would require $311032 * 311032 * 4\text{bytes} = 386\text{GB}$ of memory. In sum, large samples of high-dimensional text embeddings vectors posed both time and memory challenges.

The Self-Organizing Map (SOM) algorithm was implemented as a clustering algorithm because of its efficiency in high-dimensional spaces (Kohonen, 1982). SOM is a neural network that projects data with high-dimensional feature vectors onto a typically 2-D space while preserving the topological features of the input data (Kohonen, 1982). For each observation within the input, a feature of dimension n is mapped to a low-dimensional map and observations with similar feature values are assigned to the same region (Lee et al., 2007). Preservation of the topological feature allows for nonlinear patterns within the data to be preserved even when mapped to a lower dimension. Thus, semantically similar observations with similar vectors will be mapped to a similar region on a low-dimensional space. Additionally, it is an unsupervised approach that does not require prior knowledge of the number of clusters within the data. SOM has been used in automated ontology construction research widely due to its high-dimensional capabilities and ease of interpretation (see. Lee et al., 2007; Shih et al., 2011; Tang & Cai, 2010).

Implemented as a clustering algorithm, SOM provided good time performance without memory issues. The cluster results were evaluated using inter- and intra-cluster measures, as seen in table. However, due to the computational bottlenecks of conventional clustering algorithms, comparative measures were not generated.

4.6 Knowledge Graph Construction

The final stage of the pipeline has aggregated the output from the previous stages into a network that is analyzable using graph theory approaches. The output of the preceding Entity-Relation Extraction stage provided the basic data structure of the final KG. The extracted entity-relations triplets (E_1, rel, E_2) where E_i is the ID number of the proto-entity within with individual semantic role NPs e_i are contained. The relation upon which two proto-entity is linked will form a directed edge between them. Each verb forms an edge, and each verb-edge has the verb as its label and a weight associated with denoting the frequency of the verb within the context of its two end nodes. As such, pairs of nodes may have multiple edges (multiedges) between them, accounting for the different verbs that link them.

Formally, the final output KG is defined as an *edge-labelled directed multigraph* $G = (V, A, s, t, \ell_s, \ell_t)$ where:

- V is the set of nodes where $E_i \in V$
- A is the set of directed edges where $rel_i \in A$
- $s\colon A \to V$ and $t\colon A \to V$ denote the source and target nodes of an edge
- ℓ_s and ℓ_t are maps of the labels for each directed edge $A(s, t)$

The extended graph notations beyond the typical $G = (V, E)$ format is because of the presence of multiedges where edges incident of the same source and target nodes hold different labels and attributes (M. Newman, 2018a). To illustrate multiedges using an example form the data, a subset of the edges between the "scientist" and "climate change" proto-entities is $A(\text{scientist}, \text{climate\_change}) = \{\text{study}, \text{understand}, \text{combat}\}$. The triplet form was represented in an edge list structure that is readable by the network analysis package `NetworkX`.

4.6.1 Network Analysis

The KGs in its final network form enabled a wealth of analysis. Firstly, the basic structures of the network were described using common network statistics, including diameter for compactness, average clustering coefficient, average path length depicting, and density for the connectedness of nodes (Section 5.1).

As this paper aims to characterize the change in climate change frames and knowledge longitudinally, considerable effort was dedicated to measuring the network structure over time. To that end, two network statistics describing connectivity were calculated. First, transitivity, also known as global clustering coefficient, was calculated as:

$$C(G) = 3\frac{\#\text{ triangles}}{\#\text{ triads}} \tag{4.3}$$

. Transitivity is a measure of the clustering effects within the entire network (M. Newman, 2018b). Thus with the addition of edges forming more triangles, transitivity will also increase. However, the increase in transitivity does not necessarily translate to a better-connected network at the global level, as highly clustered subgraphs can also increase the transitivity. Thus to assess whether the increase in transitivity results from increasing local clustering effects or a global increase in network robustness, I also calculated the k-Edge Connectivity as a measure of network robustness. It measures the minimum number of k edges that can be removed before the network becomes disconnected (Kolaczyk, 2009). The two metrics elucidate whether the inclusion of new edges results in a more robust KG network or reinforces existing relations between entities. Hence, a KG with a more robust systems frame should display a similar trend to the two measures.

To operationalize the measurement of the system frame within the KGs, a measure of *Edge Robustness* (ϵ) is proposed as the ratio of k-edge connectivity and transitivity.

$$\epsilon(G) = \frac{k}{C(G)} \tag{4.4}$$

This simple ratio returns a positive number depicting the robustness of the network. A larger ratio indicates a more robust network as it is highly transitive, indicating many entities within the KG are well-connected locally, and there is also a large number of global links between what may be distant entities on a less robust network.

Thus a higher ratio indicates a more systemic coverage of climate change where links between entities are well-linked locally and globally. This ratio is especially useful for the longitudinal understanding of a KG's growth.

Lastly, a range of node centrality measures was calculated to understand the prevalence of concepts within a country's media coverage. These measures include closeness, eigenvector centrality, and communicability centrality. Closeness centrality (CC) is the mean distance of the shortest path that passes through a node. Eigenvector centrality (EC) is based on the importance of a node's adjacent nodes, where a more central node is connected to similarly highly-central nodes. These approaches are valuable in identifying the most central elements measured by the shortest path length and closest neighbours. However, on a knowledge graph, the most conveniently accessed concept may only reflect one component of frame recollection. Thus, this motivated the inclusion of communicability centrality. Communicability centrality (ComC) offers another perspective based on walks, but it also values longer walks in addition to shortest walks (Benzi & Klymko, 2013). This can accurately reflect concepts deeply embedded within a country's KG because the central node is one that many paths will pass regardless of distance. Thus, these node centrality measures will provide different perspectives on the salience of concepts within a country.

All things connected.

5 Results

Contents

5.1 Descriptive Results

The reconstruction of media frames and structural knowledge differences using KGC approaches lead to five country-specific KGs. The descriptive statistics in

Table 5.1: Global network statistics of the generated KG networks encompassing the entire analyzed time period between 1990 and 2020.

	Australia	Canada	India	UK	US
Total Nodes	118	118	118	118	118
Total Edges	6,684	10,946	5,344	7,694	10,361
Exclusive Edges	154	884	92	171	579
Diameter	2	2	3	2	2
Density	0.98	1.59	0.78	1.12	1.51
Cluster Coefficient	0.60	0.82	0.57	0.65	0.80
Avg. Path Length	1.32	1.08	1.44	1.25	1.11

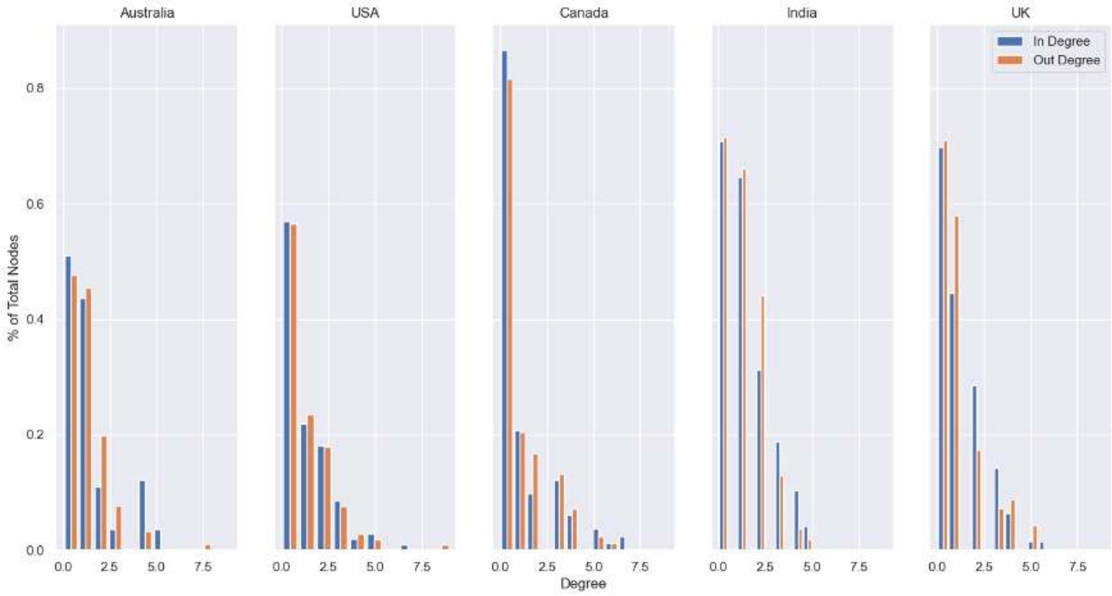


Figure 5.1: Degree distribution of the KGs of countries grouped as in and out-degrees. All countries demonstrated a power-law like degree distribution.

table 5.1show the structural differences between the complete climate change KGs for each country. The KGs share the same number of nodes because the entity clustering approach assigned all extracted arguments to a larger proto-entity cluster. (see Sec. 4.4). However, the number of edges between each KG varies significantly, with Canada and the US having over 10,000 edges or relations extracted. The measure of *Exclusive Edges* shows the number of edges not found in the union of all other networks. Canada and the US again have the most exclusive edges. However, all countries showed the presence of unique edges.

Measures of network size show that most KGs have a diameter of 2, denoting a small and highly dense network given the number of nodes. India has a slightly larger diameter of 3, translating to a less connected network. The density of the KGs, calculated as a directed graph, provides an additional metric showing the high density of most country's KGs. The presence of density values greater than 1 is due to the inclusion of self-loops. Further illuminating the density of the graphs using the average clustering coefficient as defined by Watts and Strogatz (1998) which disregards directed edges, all KGs appear to be highly clustered, with Canada having the highest average clustering coefficient. Canada, the UK, and the USA remain the densest across network metrics, demonstrating a highly connected KG.

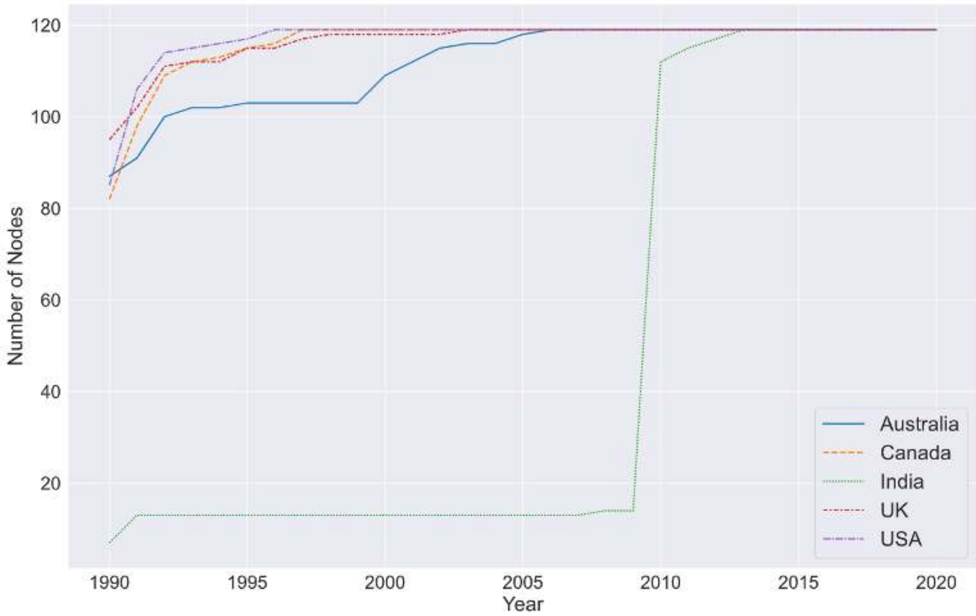


Figure 5.2: Annual count of nodes by country.

Figure 5.1 presents the in and out-degree distributions of entities within each KG. At an aggregate level, the distributions show that a power-law-like distribution is present for all countries' in and out degrees. However, Canada also saw the greatest proportion of low-degree entities than any other country, indicating many sparsely connected concepts. There are also more variations at a granular level within the distribution of lower-degree nodes of each country. For instance, the lower distribution bins of the UK and India contained more low out-degree nodes than low in-degree nodes. By contrast, Canada saw more low in-degree nodes, and the US generally has a comparable distribution of in and out-degrees throughout the degrees distributions.

5.2 Longitudinal Results

Despite each country having the same number of nodes, each took a different path to reach the full numbers of clusters, with some stabilizing significantly earlier. As seen in figure 5.2, Canada, the US, and the UK follow a similar trajectory in introducing new nodes and reach a saturated state in the 1990s. During the same period, Australia stagnated in introducing new nodes, only reaching saturation in

2006. Most interestingly is India's rapid rise in node counts in 2009, which saw the appearance of every node in the span of four years.

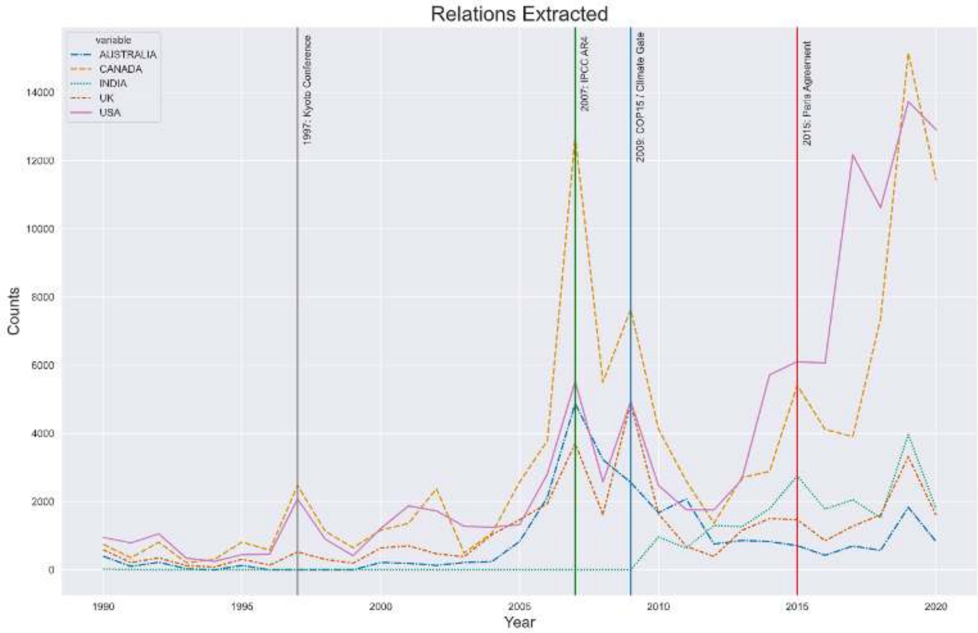
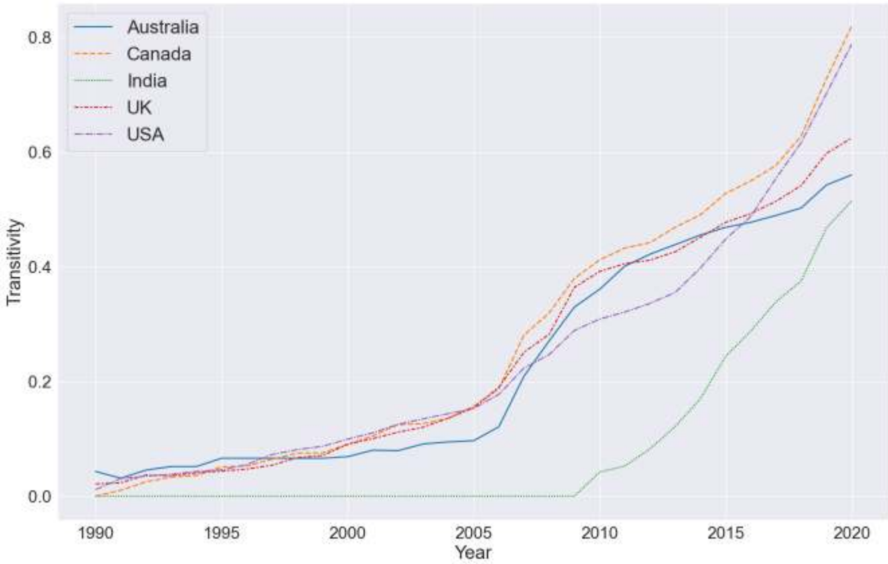


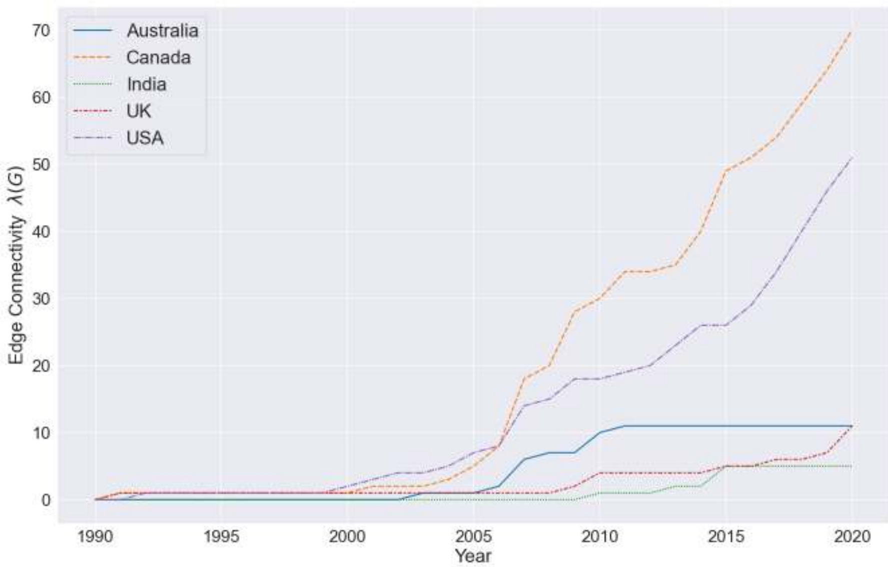
Figure 5.3: Total relations extracted per year by countries. Several defined peaks can be attributed to significant climate events demonstrating the event-driven issue-attention observation seen in literature.

Figure 5.3 shows the annual count of entity-relations extracted from the articles. There are clearly defined spikes in the relations extracted during certain years, which can be attributed to significant international climate events. Other than international events, there are country-level spikes when such patterns are not found in the other countries. For example, only Canada saw a spike in relations extracted in 2002. While most country's counts modulated congruently, there were several contrasting patterns worth noting. For example, India's rise in relations from 2009 to 2013 contrasts with the general decrease in relations from other countries during the same period. Following the spikes in relations count in 2009, Australia's count did not significantly increase while the other countries exhibited an upward trend.

Figure shows the results for measuring the connected robustness of networks as measured by transitivity and k-edge connectivity5.4. In general, the transitivity of KGs showed an increasing trend ending up highly transitive by 2020. What is most interesting in this figure is the strikingly different trend of edge connectivity



(a) Tansitivity



(b) Edge Connectivity

Figure 5.4: The longitudinal trend of the global connectivity of each country's KG. Results of two measures of network connectivity, transitivity ($C(G)$) or global clustering coefficient and k-Edge Connectivity ($\lambda(G)$) are presented.

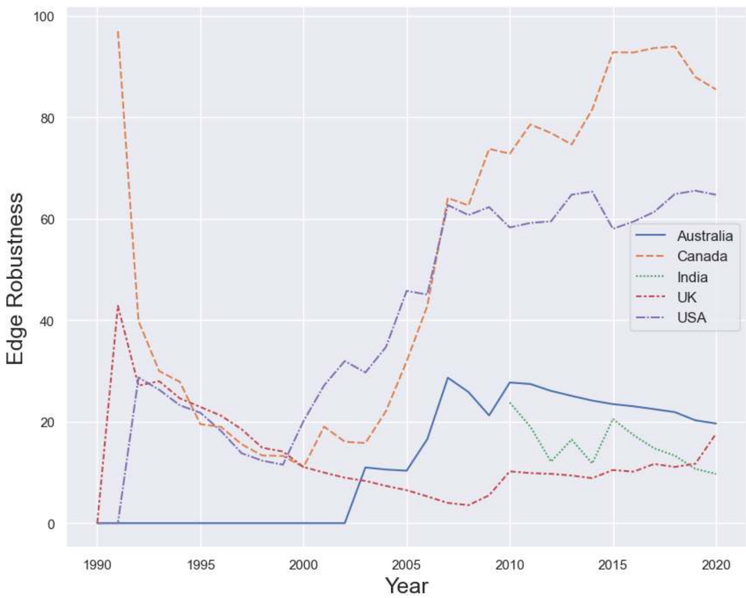


Figure 5.5: Edge robustness ϵ trends of each country's KG. Edge robustness of 0 indicates a disconnected graph as k is negative. India's trend only started in 2010 because of zero transitivity before then.

in subfigure 5.4b. The edge connectivity of Canada and the US showed a steady trend of increased robustness over the time period but rising most significantly from 2007 onward. By contrast, the edge connectivity of Australia, UK, and Indian KGs stagnated from 2010 despite its continual growth of transitivity. The most striking relationship between transitivity and edge connectivity is the UK, which saw significant growth in transitivity while edge connectivity remained.

Combining the two trends as a single measure, figure 5.5 shows the relationship between transitivity and connectivity using edge robustness. The KGs of Canada, the UK, and the USA showed high edge robustness at the beginning of the analytical period but decreased rapidly until 1999. From 2000 onward, Canada and the US generally saw an increasing trend in robustness, with Canada exceeding the US post-2007. The UK continued the earlier decreasing trend, albeit slower until 2009, which saw a modest upward trend. Australia did not begin exhibiting a robust KG until after 2003. After 2007, its robustness continued to decrease steadily. India similarly saw a generally decreasing trend in robustness. At large, it shows that Canada's introduction of new edges further reinforced the KG with increased

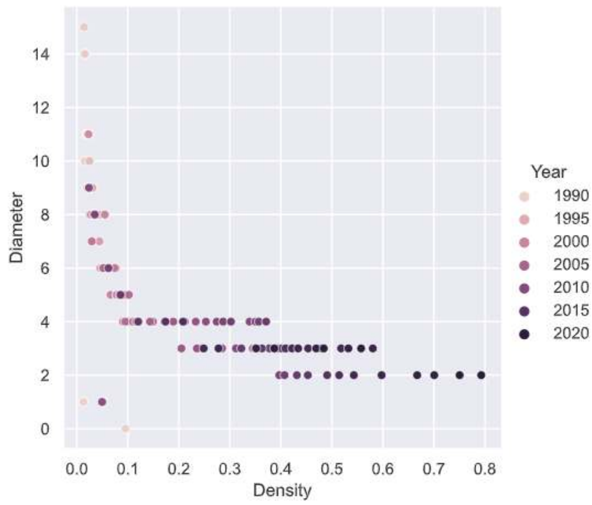


Figure 5.6: Scatterplot showing the densifying trend of where the diameter is decreasing and density increasing.

edge robustness after 2007. The other countries that remained steady or decreased steadily meant the growth in their KG did not robustly reinforce existing nodes.

The climate change KGs also showed a trend of "densification." Figure 5.6 shows that the KGs diameter decreases over time and its density correspondingly increases. The indicators are seen in the bottom left of the plot result of the diameters calculated from disconnected graphs during the early stages of a country's KG.

The longitudinal results showed that each of the KGs developed over time by including new nodes and edges, leading to changes in network structures. Although the growth led to a denser network, new edges did not contribute similarly across the countries. In Australia, UK and India, new edges did not improve the robustness of the network, while in Canada and the US, it contributed to higher edge robustness.

5.3 Centrality Results

Figure 5.7 shows the relationships of centrality measures between the top 10 nodes within each KG. Each country lies on its own level on the y-axis due to closeness centrality being affected by the total possible shortest path. Nonetheless, the relationship between CC and EC showed a near-linear relationship demonstrating a positive correlation which is confirmed by the Pearson correlation results in table 5.2.

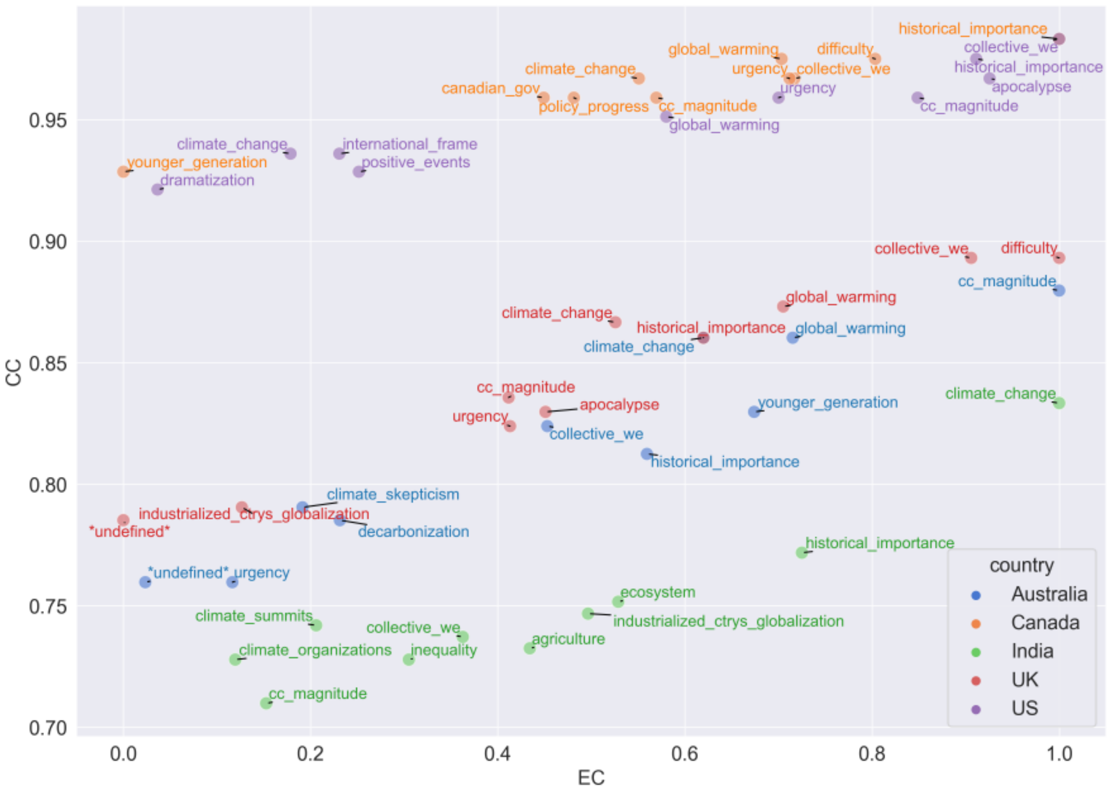


Figure 5.7: Scatterplot showing the relationships between different centrality measures for the top 10 entities ranked by eigenvector centrality. The y-axis shows the closeness centrality (CC), and the x-axis is the eigenvector centrality (EC). Note that EC is normalized between 0 and 1 for cross-country comparison.

Table 5.2: The Pearson correlation coefficients between the centrality measures. All coefficients have a $p \leq 0.001$

Corr	ComC	EC	CC
ComC	1.00	-0.37	-0.59
EC	-0.40	1.00	0.88
CC	-0.59	0.88	1.00

In Canada, the US, and Australia, the frame "collective_we," which depicts an entity cluster containing references to the collective first persons, ranks highly in EC cross-nationally ($EC > 0.7$). Within country comparison also shows their relative high ranking being second for Canada and Australia and the most central node for the US.

The nodes for "climate_change" and "global_warming" had a high rank in CC and EC for most countries, except India and the US. In the US, "climate_change" is ranked considerably less central than any other country. In India, "global_warming" does not appear within the top 10 most central nodes, but "climate_change" sees a significantly higher CC and EC than the next highest node.

The measure of communicability showed a different set of influential nodes as measured by the ability to affect all other nodes (Estrada, 2011b). Its negative correlation with CC and EC shows a considerable difference between the centrality values and ranking of nodes. Table 5.3 shows the top 10 most influential nodes for each country. Comparing countries with high article count, Canada has a higher average total communicability than the US. In the countries with fewer articles, the UK showed the highest average communicability. Within-country variations show a greater difference between the top entities.

Most countries had several unique top nodes representing specific domestic issues. Beyond those, there are greater variations among the top nodes. India and the UK are the only two countries that described industrialized countries and globalization among the top nodes. The UK is also the only country with the "historical_important" entity within the top 10. The most drastic differences in composition and ranking are seen with results in the US.

The US is the only country with climate change as the most central concept, while the next highest is Canada and India, with it in 4th place. In the US, climate change is tied centrally with urgency, collective action, international frame, climate organizations, and male public figures entities. The only other country for which climate organizations are ranked highly in India. The US also is the only country with entities describing positive events and scientists within the top 10 nodes.

While there is significant variability in each country's composition and ranking of top entities using different centrality measures, several central entities are shared across the countries even on different measures. Besides obvious top entities representing climate change and global warming, the node for collective action (collective_we) is relatively central for all countries. Among Australia, Canada, UK, and the US, nodes representing politicians and public figures are also highly central, while climate organizations are the top political entity in India. Another central frame within Australia, Canada, and the US is the legal frame entity representing

Table 5.3: Most central entities measured by communicability centrality. The values of communicability are naturally normalized and are cross-nationally comparable. ComC was calculated on an undirected graph.

	Australia		Canada		India		UK		USA	
	Avg.	0.49	Avg.	0.60	Avg.	0.43	Avg.	0.52	Avg.	0.58
	σ	0.15	σ	0.06	σ	0.19	σ	0.13	σ	0.08
	Entity	*ComC*	*Entity*	*ComC*	*Entity*	*ComC*	*Entity*	*ComC*	*Entity*	*ComC*
1	collective_we	0.723	difficulty	0.654	indian_issues	0.737	politicians	0.707	climate_change	0.662
2	younger_generation	0.721	canadian_gov	0.654	collective_we	0.729	cc_magnitude	0.699	urgency	0.662
3	cc_magnitude	0.721	male_public_figures	0.654	climate_organizations	0.720	collective_we	0.696	collective_we	0.662
4	male_public figures	0.715	climate_change	0.654	climate_change	0.715	industrialized countries and globalization	0.690	international_frame	0.662
5	politicians	0.709	legal_frame	0.654	inequality	0.711	male_public_figures	0.688	climate organizations	0.662
6	the_past	0.707	collective_we	0.654	the_past	0.707	climate_change	0.685	male_public_figures	0.662
7	climate_change	0.692	female_public figures	0.650	industrialized countries and globalization	0.707	historical_importance	0.679	positive_events	0.658
8	global_warming	0.692	younger_generation	0.650	legal_frame	0.706	international_frame	0.679	scientist	0.658
9	urgency	0.681	cc_magnitude	0.650	younger_generation	0.695	the_past	0.675	historical importance	0.658
10	legal_frame	0.674	international_frame	0.650	cc_magnitude	0.693	urgency	0.674	legal_frame	0.658

references to climate legislation. The centrality measures have revealed the most central nodes within each nation using the complete KG.

5.3.1 Longitudinal Centrality Results

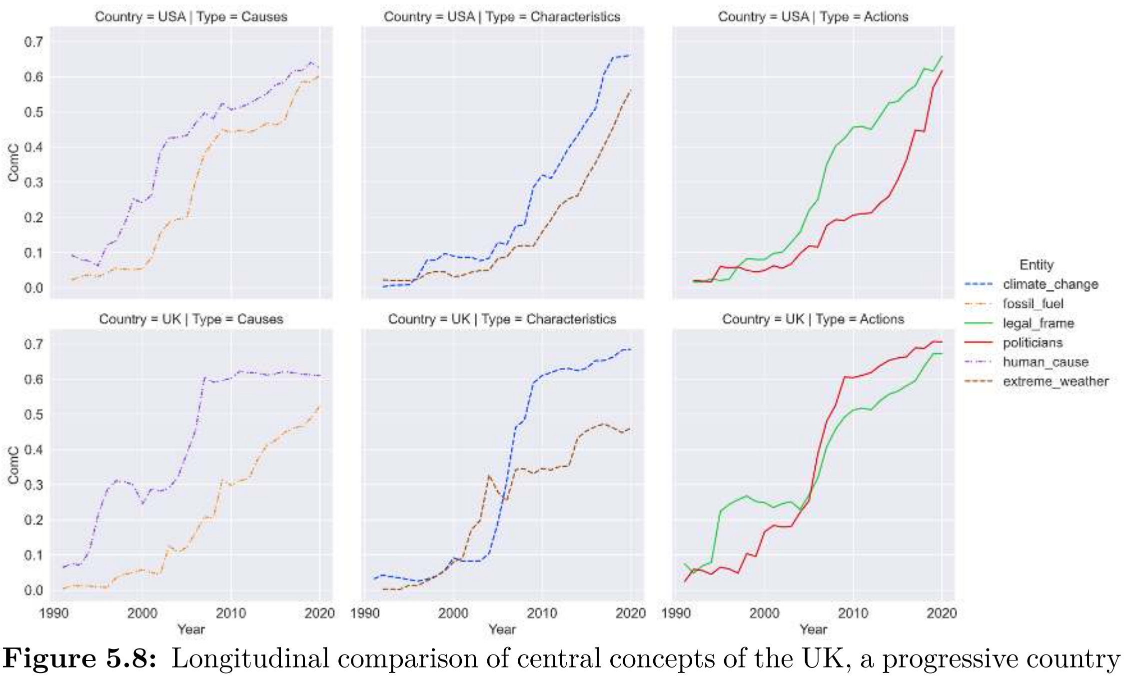


Figure 5.8: Longitudinal comparison of central concepts of the UK, a progressive country on climate change, with a less progressive country, the USA. For results of all countries, see appendix A.

This subsection presents the results of the longitudinal trends of centrality measured by communicability centrality (ComC). The plots on the figure 5.8group the trends of six representative sample entities from three concept types representing general aspects of climate change communication – the causality, climate change characteristics, and climate actions frame – to answer H1b. Two countries with considerably different progress on climate change were chosen to provide contrast. The results show a noticeable difference between the growth trends of each concept group across the two countries.

In the UK, a progressive country on climate change, a concerted sharp increase in centrality can be seen for all three concept types in 2009. The entities for climate change and human causes were especially synchronized in their increase with the Actions entities and remained highly central by plateauing. On the contrary, the US shows a more relaxed growth pattern in centrality across the three types. Climate change causality saw the earliest rise in centrality post-2000, while climate change characteristics did not grow more central until after 2010. An interesting observation is also the decoupling within the Actions frame between the politicians and legal entities. In sum, the two countries at varying stages of climate change commitments showed a drastically different trend in centrality growth between different aspects of climate change.

Be Friends of the Earth.

6
Discussion and Conclusion

Contents

6.1 Discussion

The purpose of this thesis was to develop and implement a methodology using unstructured data and a network analysis approach to identify cross-country variations in structural and longitudinal properties of climate change communication and knowledge, leading to improvements in climate change framing. This section discusses the findings from applying the methodology on climate change communications in Australia, Canada, India, the UK, and the USA. The findings are interpreted within additional contexts, and their relevant implications are presented. Lastly, this section will discuss the performance of the methodology and opportunities for future work.

The research questions raised in this thesis were:

RQ1: What are distinguishable and similar characteristics between the climate change knowledge graphs of the analyzed countries?

RQ2: What are the most salient frames within each country as measured by relational properties? How do these change over time?

The findings from analysis of the constructed KGs show that the last 30 years of climate change coverage varied drastically across countries both in the growth of network structures and their composition. Several key characteristics stand out: a) the overall network topology showed that the KGs are similarly highly connected and dense networks; b) each country's KG developed differently in structure and robustness overtime despite similar issue-attention through basic counting; c) salient concepts, as measure by centrality measures, revealed more differences in frame composition than similarities among the countries; d) individual countries followed different growth patterns over time in its central concept, progressive countries showed greater coherence in the growth of central concepts while less progressive countries developed less uniformly across concepts. These factors represent a first look at the relational properties of climate change knowledge graphs among five countries at various stages of climate change commitment.

6.2 Comparison of Climate Change KGs

The overall network statistics for each network showed the greatest amount of similarity among the countries (see sec. 5.1). The descriptive statistics presented in table 5.1show that the KGs can be considered to display "small-world" properties with a small diameter, average path length, and high density and clustering coefficient (Watts & Strogatz, 1998). Degree distribution also showed that the KGs generally follow a roughly power-law distribution found in many other empirical networks (Amaral & Scala, 2000). This entails that there are a large proportion of weakly connected concepts in each country and an increasingly smaller proportion of more connected concepts. Hence in many ways, the climate changes KGs exhibit predictable and similar qualities when measured as a static network. But, of course, media coverage is far from static, which motivated longitudinal analysis from a basic counting approach to a more complex network analysis approach.

Past research has widely studied Downs's 1972 theory on the issue-attention cycle and found it to be a common property in climate change coverage in many nations (Brossard et al., 2004; McComas & Shanahan, 1999; Schäfer et al., 2014; Schmidt et al., 2013). The approach taken to identify issue-attention diverges from past uses of article or word counts but instead uses the count of entity-relations. Despite using a different counting approach, the findings presented supports the presence of an issue-attention cycle in climate change reporting in the five analyzed countries corroborating with past research (fig. 3.1). When shown against international climate policy events, the spikes in relations-based issue-attention show a clear event-driven behaviour which is in line with the findings from Schäfer et al. However, the close coupling of issue attention frayed after 2015, which saw the US and Canada increasing in counts greatly. This may be an artifact of the imbalanced dataset, but it is nonetheless a drastic change given that in the preceding period of 2005 to 2010, Australia and the UK had comparable attention with US media.

Summarized as a partial answer to **RQ1**, it can be said that the five nations similarly share the presence of an often coupled issue-attention cycle and its possible drive by international climate events. Furthermore, the static knowledge graphs of each country are a similarly highly dense "small-world" network with a large number of paths, albeit some unique ones, connecting many concepts. The most significant difference using basic structure is highly nuanced and is better illustrated by using longitudinal structure change measured by network statics.

6.3 Longitudinal Structure Change

All the constructed KGs grew and evolved in structure over the studied time period. Growth describes the inclusion of new nodes, but more so through introducing new edges in KGs given the node creation methods (see Section 4.4). This growth pattern where nodes stabilized relatively early with growth being more edge drive led to densification and shrinking diameter. Network densification is a pattern in network growth is a well-documented pattern in network literature and real-life networks (Leskovec et al., 2007; Panzarasa et al., 2009). Specifically, the densifying pattern denotes a power-law pattern and shrinking diameter, which was formalized by Leskovec et al. in the *Forest Fire Model*.

The *Forest Fire Model* offers a theoretical explanation to this pattern of densification and shrinking diameter (Leskovec et al., 2007). As a network grows in nodes and edges, nodes have a preferential attachment to other nodes within a community, creating a denser network through communities while the overall diameter remains stable. Occasionally a new node will introduce a disproportionately large amount of out-links which "serve as “bridges” that connect formerly disparate parts of the network, bringing the diameter down" (Leskovec et al., 2007, p. 8). This can be understood as the occasional introduction of a highly reported concept that connects formerly disconnected topics in the climate change communication context. For example, these could be during releases of IPCC Assessment Reports that cover various climate science topics from natural disasters to human impacts. The exact cause of densification would require further investigation, but this provides the first evidence that some climate change KGs show characteristics of the Forest Fire Model.

Further longitudinal analysis revealed greater dissimilarities in the growth patterns. While all KGs showed the trend of densification as shown by the shrinking diameter and increasing diameter (fig. 5.6). The results specifically investigating edge connectivity and transitivity shows a less uniform growth pattern (fig. 5.4). Despite significant increases in transitivity in Australia, India, and the UK that follows a similar trend with Canada and the US, its edge connectivity had little variance after 2010. This can be understood as introducing new relations that reinforce existing clusters more often than forming new connections between disparate concepts. This is most pronounced when measured using the proposed edge robustness measure on figure 5.5, where the US and Canada continued to grow in robustness, forming more novel connections between concepts than Australia, India and UK.

The edge robustness trend must be interpreted with caution because of the data imbalance between the high and low robustness countries. In Australia, UK, and India, only one publication was included due to data extraction limitations. The data balance caveat may actually inadvertently provide evidence on the effects of ideologically driven framing. Carvalho (2007) argued that individual media organizations reflect their own "ideological culture" within their coverage, which is reflected through the selection and framing of scientific news. This argument is in line with the high levels of edge robustness observed in countries with two publication sources compared to the low levels in single publication countries. The observation of considerably higher edge robustness in countries with two publications than those with one is likely a manifestation of publication-specific ideologies in coverage. Put differently, an individual may experience an "echo chamber" effect if only a single media is consumed where existing frames and relations are continuously reinforced; whereas, consuming coverage from multiple sources introduces more novel connections (Dubois & Blank, 2018). An implication tangential to the research questions from this finding is empirically demonstrating the effects of single-source media consumption and its possible echo chamber effects.

At this point, we can provide more answers to **RQ1** on the similarities and differences between the KGs. All knowledge graphs showed a densification trend over time, although the source of densification is not uniform. In countries with two media analyzed, relations between more disparate concept clusters were more prominent and likely responsible for densifying than countries with only a single media analyzed. Canada showed a considerably more robust KG than the US, even between countries with two media, which indicates more diversely connected concept clusters.

6.4 Salient Frames

This section provides the answer to **RQ2** on identifying salient frames in the analyzed countries and their longitudinal patterns. Prior literature has emphasized the importance of frames for communicating climate change, especially in shaping the mental model to understand climate change (Bostrom et al., 1994; Moser, 2010;

Spence & Pidgeon, 2010). However, there was very little large-scale cross-country longitudinal analysis of frames for the exceptions of some two-country comparative frame studies (see section). Furthermore, informed by the relational property of understanding raised by the theory of frame semantics, this research question sought to identify salient climate change frames using network centrality measures over time.

Two sets of centrality measures were calculated for each KG, closeness and eigenvector centrality were calculated on the directed graph, and communicability centrality was calculated on the underacted graph to provide multiple perspectives on the ranking of concepts. Interpreting centrality rankings is network and domain-specific, and their interpretations have been formally discussed widely in social network analysis but to a significantly less extent for semantic network and knowledge graph networks. Therefore, interpretations made on ranks will use interpretations borrowed from social network analysis and re-interpreted in semantic networks.

6.4.1 Immediately Salient Frames

Central concepts were first identified using the static full KGs. The full KGs can be understood as a snapshot of all media coverage in the five countries from 1990 to 2020. The most identifiable finding from the scatterplot of central concepts on figure 5.7 is the linear-like relationship between closeness and eigenvector centrality. A node is central in closeness if it has a small mean distance with all other nodes (M. Newman, 2018b). An eigenvector central node is connected to many similarly highly central nodes (M. Newman, 2018b). Therefore, the linear relationship from the results shows that the highly central concepts in all the KGs are not only close in the distance to other nodes but also connected to similarly highly salient concepts.

In the context of semantic networks, ranking by closeness and eigenvector centrality provides an interpretation for concept salience. A node high in closeness centrality takes few mental connections to arrive at from any other concept on the KG. On the other hand, eigenvector centrality is a measure of influence based on the centrality of a concept's neighbours, which means a concept high in eigenvector centrality can be understood as an influential concept in the KG. An important consideration to the interpretation of these two measures is their mathematical foundation, which uses the shortest-path length for closeness and the adjacent neighbours of a node for eigenvector centrality (see M. Newman (2018b)). The use of such elements emphasizes immediacy such that the most salient nodes measured by closeness and eigenvector centrality require the least mental connections.

Under the interpretation of the two measures as the most immediate and influential relative frames, the implications of the centrality results are interpreted here. The concepts of collective action, the historical importance of climate change, apocalyptic consequences, and the magnitude of climate change are highly clustered in the US. They are easily evoked simultaneously when thinking about climate change, whereas actual concepts related to climate change are the least likely to be evoked out of all countries. The prevalence of the apocalyptic frame in the US is similarly reported by Foust and O'Shannon Murphy (2009), but its position as a highly salient and influential concept is a concerning finding. Apocalyptic and fear framing have generally been shown to be ineffective frames that often "backfire "(Feinberg & Willer, 2011; Witte, 1994). Contrasted with results from other countries, the apocalyptic frame was not observed as a top concept.

Somewhat reassuring is the high salience of the collective action frame across countries, indicating a considerable amount of coverage linking climate change as a collective issue. Past research quantifying the salience of the collective frame is scarce, but the current finding for the US contradicts one such study. In Olausson (2009), the authors suggested that the collective action frame is less salient in US media as a result of false-balance, but the present result does not support such hypothesis. Although, the current methodological approach does not account for the sentiment of the individual concepts that the collective action entity is built upon. Nonetheless, the quantification of the collective actions frame opens up opportunities for future work to confirm or contrast the preliminary results showing its widespread prevalence presented here.

6.4.2 Embedded Frames

Using communicability betweenness, which measures centrality based on all possible paths that pass through a node, provides an alternative perspective on the importance ranking of concepts. The communicability of a concept can be understood as its embeddedness within a knowledge graph. An embedded concept can be reached from another concept in several short or many long mental steps or a combination of the two (Estrada, 2011b). In short, it is a concept that all other concepts eventually lead to.

The communicability centrality results showed a different set of top concepts, including those missing from the ranking by the previous measures. Across all countries, entities related to politicians and public figures became highly ranked (table 5.3). The politician and public figure entities may be partially explained by the "celebrification" trend within climate change coverage by **boykoff_Conspicuous_2009**. The authors found evidence that the news media cite a growing number of highly politicized celebrities and famous political actors. Another explanation to this phenomenon is offered by Carvalho (2007, p.232), which stated that "governmental moves to control and recontextualize understandings of the greenhouse effect led to most media discourse being taken over by politicians and other actors". Under the assumption that public figures cited are more often advocates for climate change actions, an alternative perspective is to view the entity clusters as agents related to climate change solutions.

By considering climate public figures, politicians and similarly highly ranked entities such as climate organizations and the legal frame as solution agents, we can begin exploring the feasibility of H2a, which states:

> **H2a**: Countries with more progress on climate change have more salient frames relevant to solutions and current policies than causes and consequences.

The UK and India are more progressive countries at 5th and 10th place on the CCPI, respectively. In the UK, we can see politicians as the most embedded entity. In India, the "indian_issues" cluster, which contains a significant amount of references to Indian public figures[1] is also the most embedded. In contrast, the embeddedness of public figures is much lower in Australia and the US. Although Canada ranks low on the CCPI, entities for Canadian government agencies and public figures were highly embedded, contradicting the hypothesis. Regarding the consequence and causal frames, results are mixed at best with concepts describing consequences and causes occurring throughout the top ranks without a clear pattern. This finding contradicts the expectations of H1a and may suggest a more nuanced agenda-setting effect between media, the public and political actors (**hilgartner_1988**). In light of the mixed evidence provided, H2a cannot be supported, and there is no clear relationship between a country's climate performance and the embeddedness of solutions or casual and consequential frames.

6.4.3 Longitudinal Frame Salience

By combining our methodology's relational and longitudinal abilities, frame salience was tracked through the analyzed periods. In this subsection, the findings in the context of H2b is discussed. The hypothesis states that:

> **H2b**: The most salient frames in countries most progressive on climate change see causal, consequential, and action frames stabilized early within the analyzed period.

Prior studies have noted the system complexity of climate change and the need for communication models that enable systems understanding (Pearce et al., 2015b; Sterman, 2011). Motivated by such a gap, H2a was proposed as one such communication model where the causal, consequential, and action frames of climate change are brought to a high level of salience and kept in the knowledge. The use of communicability centrality was deemed the most appropriate measure for both salience and embeddedness based on its calculation using all possible paths and path lengths (Benzi & Klymko, 2013; Estrada, 2011a).

In the current study, comparing a climate progressive country (UK) with a lesser one (USA) showed that the embeddedness and salience of the causal, consequential

[1] As a limitation of the clustering model, Indian names and titles were poorly clustered with Indian words and cities hence the broad cluster name.

and actions frame increased in a sharp and concerted manner. In the UK, the three frames rose sharply from 2005 until 2008 (fig. 5.8), which is the year that the UK passed the Climate Change Act 2008 (2008) — the world's first legally-binding net-zero target. The bill was a direct result of the 18-month Big Ask campaign led by the Friends of the Earth, an environmental non-governmental organization (of the Earth, 2017). The campaign exploited the peak of the attention cycle and launched a high-profile campaign that gripped the attention of the UK from individuals to business and political figures (Carter & Childs, 2018). The sustained media attention drove cross-party and nationwide support for a climate bill (Carter & Childs, 2018). The findings presented in figure 5.8 further illustrate the campaign's impact by showing that the embeddedness of the causal, characteristic and action frame was tightly coupled during its sharp rise. By comparison, the US nor any other country[2] displayed such a simultaneous rise to high levels of embeddedness.

Therefore, the findings suggest that a "media blitz" style of communication model where all three aspects of climate change are covered in a highly connected and salient manner is an effective potential communication model to establishing the momentum necessary to achieving political climate action. Moreover, to suggest somewhat speculatively, such an approach may be effective at constructing systems frame as suggested in Sterman (2011) by simultaneously connecting multiple dimensions of climate change. Which, as the Big Ask campaign has shown, is possible. Hypothesis H2b, therefore, is supported by the evidence. In addition, findings showed that the causal, characteristic and action frame stabilized early and rose sharply to high levels of embeddedness in a cohesive manner.

6.5 Conclusion

The purpose of this paper was to develop a methodology to analyze past climate change communications using unstructured texts empirically. Despite the exploratory nature, this study offers some insights into the relationships between framing, knowledge formation, and climate progress. The methodology drew on

[2]Except perhaps India, but the available period of evidence is too short to be considered.

knowledge graph construction and open information extraction techniques to enable a network analysis of cross-country and longitudinal patterns in climate change framing. This study has shown that despite the coupling of issue-attention cycles, knowledge growth is not uniform. Some countries and the use of a single media source exhibit an "echo-chamber" effect where highly connected concept clusters become reinforced over time. Furthermore, this study showed that the growth pattern of salient frames in more climate progressive countries is more cohesive and tightly coupled. An implication of this is the recommendation of a communication approach that can effectively construct the system framing of climate change.

The generalisability of these results is subject to at least two limitations. Firstly, data availability led to three countries having only a single publication analyzed. This was addressed by separating some interpretations of results from countries with more than one publication. Secondly, the knowledge graph construction pipeline can be improved by including co-reference resolution and entity recognition. However, as shown in this study, the present pipeline is a minimally viable approach. Notwithstanding the limitations, the knowledge graph approach of content analysis has opened up opportunities for applying more empirical network analytic such as the quadratic assignment procedure to elucidate relationships of content structure with socioeconomic variables or other countries.

Appendices

A

Supplementary Data

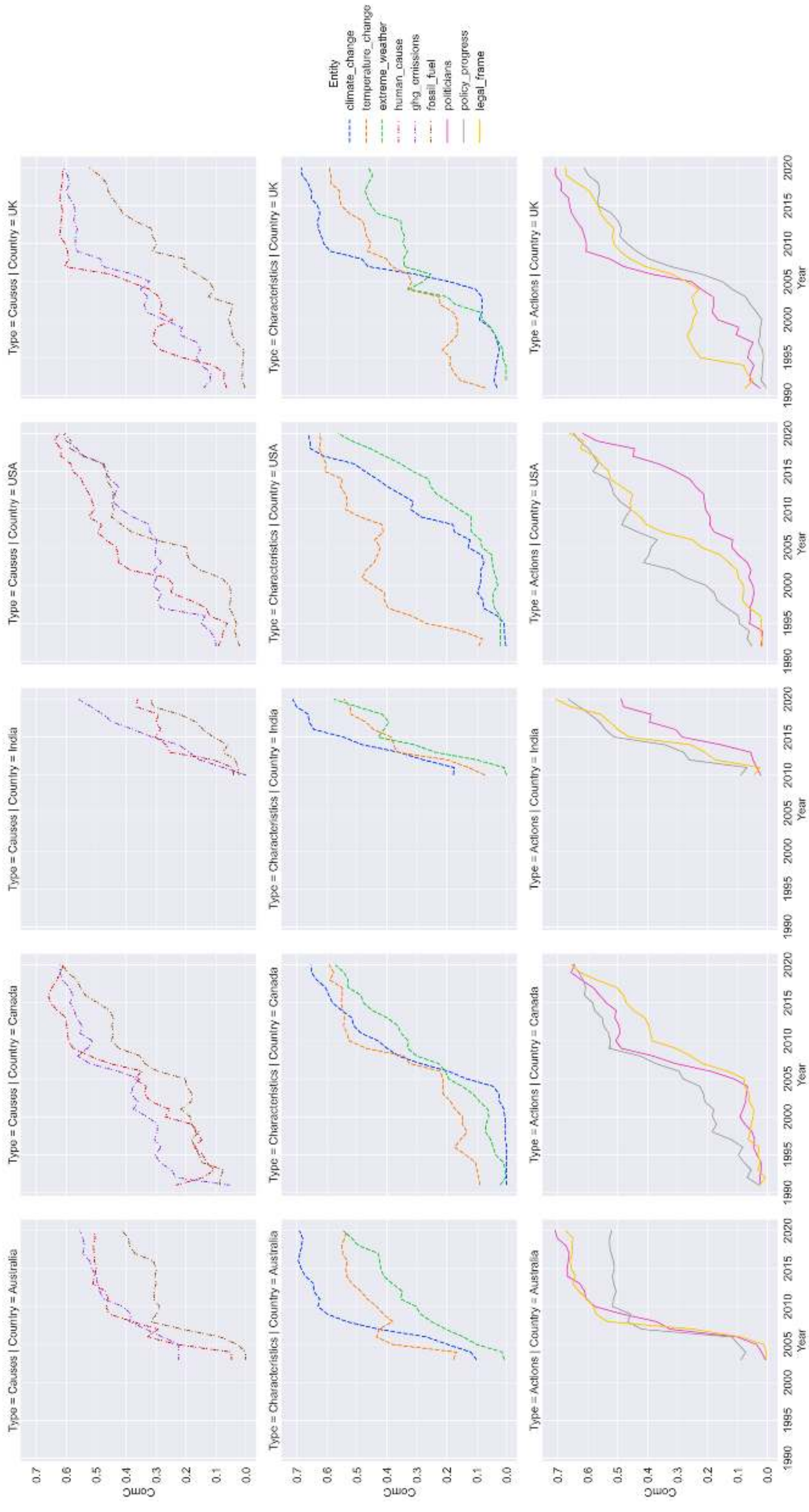
Contents

A.1 Annual Network Statistics

Table A.1: Annual network statistics.

Measure	Country	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Avg. Clustering	Australia	0.03	0.03	0.05	0.06	0.06	0.07	0.07	0.07	0.07	0.07	0.07	0.08	0.07	0.09	0.09	0.11	0.13	0.35	0.29	0.22	0.43	0.38	0.45	0.47	0.49	0.50	0.51	0.52	0.54	0.58	0.60
	Canada	0.00	0.00	0.02	0.02	0.03	0.04	0.05	0.07	0.08	0.08	0.10	0.11	0.13	0.13	0.14	0.16	0.19	0.38	0.32	0.28	0.43	0.41	0.44	0.47	0.49	0.53	0.55	0.58	0.63	0.73	0.82
	India	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.04	0.10	0.13	0.19	0.27	0.32	0.37	0.41	0.52	0.57
	UK	0.01	0.01	0.03	0.03	0.03	0.05	0.05	0.05	0.06	0.07	0.09	0.10	0.11	0.11	0.13	0.15	0.19	0.37	0.29	0.25	0.42	0.40	0.43	0.44	0.47	0.49	0.51	0.53	0.56	0.62	0.65
	USA	0.01	0.02	0.04	0.04	0.04	0.05	0.06	0.08	0.09	0.09	0.11	0.12	0.13	0.14	0.15	0.16	0.18	0.29	0.25	0.23	0.32	0.31	0.34	0.36	0.40	0.45	0.49	0.56	0.62	0.71	0.80
Avg. Degree	Australia	31.97	32.13	31.72	32.00	32.00	32.34	32.34	32.34	32.34	32.34	33.15	33.73	33.96	34.11	34.62	36.62	40.22	63.72	57.67	50.53	71.30	66.92	73.11	74.56	76.15	77.45	78.26	79.79	81.16	85.29	87.05
	Canada	31.60	31.31	32.38	32.26	32.32	33.10	33.58	35.57	36.50	37.07	38.12	39.37	41.11	41.54	42.35	44.60	47.89	70.78	63.56	58.79	76.83	74.46	78.05	80.89	83.39	87.82	90.57	93.94	100.31	112.39	123.17
	India	37.21	33.46	33.46	33.46	33.46	33.46	33.46	33.46	33.46	33.46	33.46	33.46	33.46	33.46	33.46	33.46	33.46	31.89	31.89	33.46	34.00	32.78	37.58	40.44	44.47	50.74	54.77	59.50	62.91	71.53	75.69
	UK	32.66	33.11	32.92	32.85	33.02	33.48	33.71	34.44	34.97	35.21	36.55	37.86	38.80	39.60	41.55	44.00	47.83	66.89	57.94	54.40	71.39	70.04	72.14	73.83	76.67	79.52	81.23	83.45	86.41	92.70	95.61
	USA	30.75	31.27	32.57	33.08	33.41	33.77	34.18	35.78	36.57	36.98	38.15	39.55	41.00	42.21	43.28	44.59	47.19	60.00	55.17	52.53	63.96	62.47	65.75	68.13	72.83	78.65	83.44	91.01	98.29	108.47	118.21
Density	Australia	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03	0.04	0.05	0.08	0.28	0.23	0.17	0.35	0.31	0.37	0.38	0.39	0.40	0.41	0.42	0.43	0.47	0.48
	Canada	0.01	0.01	0.02	0.02	0.02	0.03	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10	0.10	0.12	0.15	0.35	0.28	0.24	0.40	0.38	0.41	0.43	0.45	0.49	0.51	0.54	0.60	0.70	0.79
	India	0.10	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.02	0.06	0.09	0.12	0.17	0.21	0.25	0.28	0.35	0.39
	UK	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.04	0.04	0.04	0.05	0.07	0.07	0.08	0.10	0.12	0.15	0.31	0.24	0.21	0.35	0.34	0.36	0.37	0.40	0.42	0.43	0.45	0.48	0.53	0.56
	USA	0.01	0.02	0.02	0.02	0.03	0.03	0.03	0.05	0.05	0.06	0.07	0.08	0.09	0.10	0.11	0.12	0.14	0.25	0.21	0.19	0.29	0.27	0.30	0.32	0.36	0.41	0.45	0.52	0.58	0.67	0.75
Diameter	Australia	10	11	11	11	11	11	11	11	11	11	8	8	7	8	8	6	5	3	4	4	3	3	3	3	3	3	3	3	3	3	3
	Canada	1	15	11	9	9	8	8	7	6	6	6	6	5	5	5	4	4	3	3	3	2	3	2	2	2	2	2	2	2	2	2
	India	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	8	9	6	5	4	4	4	3	3	3	3
	UK	14	14	11	11	11	9	9	8	8	8	8	6	6	5	4	4	4	3	3	3	4	4	4	4	3	3	3	3	3	3	3
	USA	1	14	11	10	9	8	8	6	6	6	5	5	4	4	4	4	4	4	4	4	4	4	4	3	3	3	3	3	3	2	2
Edges	Australia	113	129	189	204	204	235	235	235	235	235	293	342	379	430	488	706	1159	3931	3218	2375	4826	4309	5040	5211	5398	5552	5647	5828	5989	6477	6684
	Canada	87	133	222	247	281	349	397	610	720	787	911	1058	1264	1314	1410	1675	2063	4765	3913	3350	5479	5199	5622	5957	6253	6775	7100	7497	8249	9674	10946
	India	4	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	9	9	8	452	290	827	1184	1660	2400	2875	3434	3836	4853	5344
	UK	135	171	269	289	308	366	392	490	561	589	746	899	1009	1085	1315	1604	2057	4306	3250	2832	4837	4677	4925	5124	5459	5796	5998	6260	6609	7351	7694
	USA	93	171	282	320	352	398	446	634	728	776	914	1079	1251	1393	1520	1674	1981	3493	2922	2611	3960	3784	4171	4452	5007	5693	6258	7152	8011	9212	10361
Nodes	Australia	86	90	99	101	101	102	102	102	102	102	108	111	114	115	115	117	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118
	Canada	82	97	108	111	112	114	115	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118
	India	7	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	14	14	13	114	111	116	118	118	118	118	118	118	118	118
	UK	93	101	110	111	111	114	114	116	117	117	117	117	117	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118
	USA	85	105	113	114	115	116	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118	118
Transitivity	Australia	0.043	0.031	0.045	0.052	0.052	0.066	0.066	0.066	0.066	0.066	0.069	0.080	0.079	0.091	0.095	0.097	0.121	0.330	0.271	0.209	0.401	0.361	0.422	0.438	0.455	0.469	0.477	0.489	0.502	0.542	0.560
	Canada	0.000	0.010	0.025	0.033	0.036	0.051	0.053	0.064	0.075	0.075	0.091	0.105	0.125	0.127	0.136	0.157	0.187	0.379	0.319	0.281	0.432	0.412	0.442	0.469	0.490	0.528	0.549	0.576	0.628	0.728	0.819
	India	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.053	0.042	0.082	0.121	0.170	0.244	0.288	0.339	0.376	0.467	0.514
	UK	0.021	0.023	0.037	0.036	0.041	0.044	0.047	0.054	0.067	0.071	0.090	0.100	0.112	0.120	0.136	0.154	0.190	0.364	0.282	0.251	0.405	0.392	0.411	0.426	0.451	0.477	0.493	0.514	0.541	0.598	0.624
	USA	0.011	0.031	0.035	0.038	0.043	0.046	0.055	0.073	0.081	0.087	0.099	0.110	0.125	0.135	0.144	0.153	0.177	0.289	0.247	0.223	0.321	0.309	0.336	0.355	0.398	0.448	0.488	0.554	0.617	0.702	0.788
Undirected Edge Connectivity	Australia	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	2	7	7	6	11	10	11	11	11	11	11	11	11	11	11
	Canada	0	1	1	1	1	1	1	1	1	1	1	2	2	2	3	5	8	28	20	18	34	30	34	35	40	49	51	54	59	64	70
	India	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	2	2	5	5	5	5	5	5
	UK	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	4	4	4	4	4	5	5	6	6	7	11
	USA	0	0	1	1	1	1	1	1	1	1	2	3	4	4	5	7	8	18	15	14	19	18	20	23	26	26	29	34	40	46	51

A.2 Full Longitudinal Frame Comparison



Works Cited

Al-Aswadi, F. N., Chan, H. Y., & Gan, K. H. (2020). Automatic ontology construction from text: A review from shallow to deep learning trend. *Artificial Intelligence Review*, *53*(6), 3901–3928.

Amaral, L. A. N., & Scala, A. (2000). Classes of small-world networks, 4.

Anderson, A. (2009). Media, politics and climate change: Towards a new research agenda [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1751-9020.2008.00188.x]. *Sociology Compass*, *3*(2), 166–182.

Antilla, L. (2005). Climate of scepticism: US newspaper coverage of the science of climate change. *Global Environmental Change*, *15*(4), 338–352.

Bach, N., & Badaskar, S. (2007). A review of relation extraction, 15.

Baker, C. F., Fillmore, C. J., & Lowe, J. B. (1998). The berkeley FrameNet project. *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*, 86–90.

Barkemeyer, R., Figge, F., Hoepner, A., Holt, D., Kraak, J. M., & Yu, P.-S. (2017). Media coverage of climate change: An international comparison [Publisher: SAGE Publications Ltd STM]. *Environment and Planning C: Politics and Space*, *35*(6), 1029–1054.

Beckert, J. (2010). How do fields change? the interrelations of institutions, networks, and cognition in the dynamics of markets [Publisher: SAGE Publications Ltd]. *Organization Studies*, *31*(5), 605–627.

Benzi, M., & Klymko, C. (2013). Total communicability as a centrality measure. *Journal of Complex Networks*, *1*(2), 124–149.

Bord, R. J., O{\textbackslash}textquotesingleConnor, R. E., & Fisher, A. (2000). In what sense does the public need to understand global climate change? [Publisher: SAGE Publications]. *Public Understanding of Science*, *9*(3), 205–218.

Bornkessel, I., Schlesewsky, M., Comrie, B., & Friederici, A. D. (2009, September 24). Introduction [Google-Books-ID: cukkO31NYEcC]. *Semantic role universals and argument linking: Theoretical, typological, and psycholinguistic perspectives*. Walter de Gruyter.

Bostrom, A., Morgan, M. G., Fischhoff, B., & Read, D. (1994). What do people know about global climate change? 1. mental models [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1539-6924.1994.tb00065.x]. *Risk Analysis*, *14*(6), 959–970.

Boykoff, M. T., & Boykoff, J. M. (2004). Balance as bias: Global warming and the US prestige press. *Global Environmental Change*, *14*(2), 125–136.

Boykoff, M. T., & Boykoff, J. M. (2007). Climate change and journalistic norms: A case-study of US mass-media coverage. *Geoforum*, *38*(6), 1190–1204.

Brossard, D., Shanahan, J., & McComas, K. (2004). Are issue-cycles culturally constructed? a comparison of french and american coverage of global climate change. *Mass Communication and Society*, *7*(3), 359–377.

Burck, J., Hagen, U., Bals, C., Höhne, N., & Nascimento, L. (2021). *Climate change performance index: Background and methodology.*

Camacho-Collados, J., & Pilehvar, M. T. (2018). On the role of text preprocessing in neural network architectures: An evaluation study on text categorization and sentiment analysis. *arXiv:1707.01780 [cs]*. Retrieved August 24, 2021, from http://arxiv.org/abs/1707.01780

Capstick, S., & Lorraine, W. (2018). Public perceptions of climate change. In B. Brevini & J. Lewis (Eds.), *Climate change and the media*. Peter Lang.

Capstick, S., Whitmarsh, L., Poortinga, W., Pidgeon, N., & Upham, P. (2015). International trends in public perceptions of climate change over the past quarter century [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.321]. *WIREs Climate Change*, *6*(1), 35–61.

Carley, K. M., & Kaufer, D. S. (1993). Semantic connectivity: An approach for analyzing symbols in semantic networks. *Communication Theory*, *3*(3), 183–213.

Carter, N., & Childs, M. (2018). Friends of the earth as a policy entrepreneur: the big ask' campaign for a UK climate change act [Publisher: Routledge _eprint: https://doi.org/10.1080/09644016.2017.1368151]. *Environmental Politics*, *27*(6), 994–1013.

Carvalho, A. (2007). Ideological cultures and media discourses on scientific knowledge: Re-reading news on climate change [Publisher: SAGE Publications Ltd]. *Public Understanding of Science*, *16*(2), 223–243.

Carvalho, A., & Burgess, J. (2005). Cultural circuits of climate change in u.k. broadsheet newspapers, 1985–2003 [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1539-6924.2005.00692.x]. *Risk Analysis*, *25*(6), 1457–1469.

Change, I. C. et al. (2007). The physical science basis.

Climate Change Act 2008 (2008, November 26). Retrieved September 9, 2021, from https://www.legislation.gov.uk/ukpga/2008/27/contents

Doerfel, M. L. (1998). What constitutes semantic network analysis? a comparison of research and methodologies. *Connections*, *21*(2), 16–26.

Downs, A. (1972). Up and down with ecology: The issue-attention cycle. [OCLC: 953969227]. *Agenda setting: Readings on media, public opinion, and policymaking* (pp. 462–473). Retrieved July 16, 2021, from https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=1288637

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*(3), 547.

Dryzek, J. S., Norgaard, R. B., & Schlosberg, D. (2011, August 18). *The oxford handbook of climate change and society* [Google-Books-ID: RsYr_iQUs6QC]. OUP Oxford.

Dubois, E., & Blank, G. (2018). The echo chamber is overstated: The moderating effect of political interest and diverse media [Publisher: Routledge]. *Information, Communication & Society*, *21*(5), 729–745.

Ehrlinger, L., & Wöß, W. (2016). Towards a definition of knowledge graphs, 4.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1460-2466.1993.tb01304.x]. *Journal of Communication*, *43*(4), 51–58.

Estrada, E. (2011a). *Centrality measures* [Publication Title: The Structure of Complex Networks Section: The Structure of Complex Networks]. Oxford University Press. Retrieved September 9, 2021, from https://ezproxy-prd.bodleian.ox.ac.uk:2196/view/10.1093/acprof:oso/9780199591756.001.0001/acprof-9780199591756-chapter-7

Estrada, E. (2011b). *Communicability functions in networks* [Publication Title: The Structure of Complex Networks Section: The Structure of Complex Networks]. Oxford University Press. Retrieved September 9, 2021, from https://ezproxy-prd.bodleian.ox.ac.uk:2196/view/10.1093/acprof:oso/9780199591756.001.0001/acprof-9780199591756-chapter-6

Farrell, J. (2019). The growth of climate change misinformation in US philanthropy: Evidence from natural language processing [Publisher: IOP Publishing]. *Environmental Research Letters*, *14*(3).

Feinberg, M., & Willer, R. (2011). Apocalypse soon?: Dire messages reduce belief in global warming by contradicting just-world beliefs [Publisher: SAGE Publications Inc]. *Psychological Science*, *22*(1), 34–38.

Fillmore, C. J. (2006, September 19). Frame semantics [Series Title: Cognitive Linguistics Research]. In D. Geeraerts (Ed.). D. Geeraerts, R. Dirven, J. R. Taylor, & R. W. Langacker (**typeredactors**), *Cognitive linguistics: Basic readings* (pp. 373–400). Mouton de Gruyter.

Foust, C. R., & O'Shannon Murphy, W. (2009). Revealing and reframing apocalyptic tragedy in global warming discourse [Publisher: Routledge _eprint: https://doi.org/10.1080/17524030902916624]. *Environmental Communication*, *3*(2), 151–167.

Grundmann, R., & Scott, M. (2014). Disputed climate science in the media: Do countries matter? *Public Understanding of Science*, *23*(2), 220–235.

He, L., Lee, K., Lewis, M., & Zettlemoyer, L. (2017). Deep semantic role labeling: What works and what's next. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 473–483.

Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., de Melo, G., Gutierrez, C., Gayo, J. E. L., Kirrane, S., Neumaier, S., Polleres, A., Navigli, R., Ngomo, A.-C. N., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., & Zimmermann, A. (2021). Knowledge graphs. *arXiv:2003.02320 [cs]*. Retrieved May 13, 2021, from http://arxiv.org/abs/2003.02320

Houghton, E. (1996). *Climate change 1995: The science of climate change: Contribution of working group i to the second assessment report of the intergovernmental panel on climate change* (Vol. 2). Cambridge University Press.

IPCC. (2001). *Climate change 2001: Synthesis report*.

IPCC. (2021). *IPCC 2021: Summary for policymakers*. Retrieved August 18, 2021, from https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_SPM.pdf

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks [Bandiera_abtest: a Cg_type: Nature Research Journals Number: 10 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Climate change;Psychology Subject_term_id: climate-change;psychology]. *Nature Climate Change*, *2*(10), 732–735.

Kang, G. J., Ewing-Nelson, S. R., Mackey, L., Schlitt, J. T., Marathe, A., Abbas, K. M., & Swarup, S. (2017). Semantic network analysis of vaccine sentiment in online social media. *Vaccine*, *35*(29), 3621–3638.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, *43*(1), 59–69.

Kolaczyk, E. D. (2009). *Statistical analysis of network data*. Springer New York.

Kopstein, F. F. (1969). Graph theory as a metalanguage of communicable knowledge, 26.

Kovecses, Z. (2006, October 12). *Language, mind, and culture: A practical introduction* [Google-Books-ID: rvgafOskC6gC]. Oxford University Press.

Lee, C.-S., Kao, Y.-F., Kuo, Y.-H., & Wang, M.-H. (2007). Automated ontology construction for unstructured text documents. *Data & Knowledge Engineering*, *60*(3), 547–566.

Leskovec, J., Kleinberg, J., & Faloutsos, C. (2007). Graph evolution: Densification and shrinking diameters. *arXiv:physics/0603229*. Retrieved September 10, 2021, from http://arxiv.org/abs/physics/0603229

Liu, J. C. E., & Zhao, B. (2017). Who speaks for climate change in china? evidence from weibo [Publisher: Climatic Change]. *Climatic Change*, *140*(3), 413–422.

Liu, Y.-c., Wu, C., & Liu, M. (2011). Research of fast SOM clustering for text information. *Expert Systems with Applications*, *38*(8), 9325–9333.

Löfstedt, R. E. (1991). Climate change perceptions and energy-use decisions in northern sweden. *Global Environmental Change*, *1*(4), 321–324.

Manzanedo, R. D., & Manning, P. (2020). COVID-19: Lessons for the climate change emergency [Publisher: Elsevier B.V.]. *Science of the Total Environment*, *742*, 140563.

Matthes, J. (2007). *Framing-Effekte: Zum Einfluss der Politikberichterstattung auf die Einstellungen der Rezipienten* (1st ed., Vol. 13). Nomos Verlagsgesellschaft mbH & Co. KG.

Matthews, P. H. (2007). *The concise oxford dictionary of linguistics* (2nd ed) [OCLC: ocn167505258]. Oxford University Press.

Maynard, D., Bontcheva, K., & Augenstein, I. (2017). *Natural language processing for the semantic web*. Morgan & Claypool Publishers.

McComas, K., & Shanahan, J. (1999). Telling stories about global climate change: Measuring the impact of narratives on issue cycles [Publisher: SAGE Publications Inc]. *Communication Research*, *26*(1), 30–57.

Moser, S. C. (2010). Communicating climate change: History, challenges, process and future directions [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.11]. *WIREs Climate Change*, *1*(1), 31–53.

Moser, S. C. (2016). Reflections on climate change communication research and practice in the second decade of the 21st century: What more is there to say? [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.403]. *WIREs Climate Change*, *7*(3), 345–369.

National Research Council. (1979, March 13). *Carbon dioxide and climate*. National Academies Press.

Newman, M. (2018a, October 18). 6. mathematics of networks [Issue: 9 ISSN: 1098-6596]. *Networks* (2nd ed., pp. 1689–1699). Oxford University Press.

Newman, M. (2018b). 7. measures and metrics. *Networks* (pp. 304–339).

Newman, N., Fletcher, R., Schulz, A., Andi, S., & Nielsen, R. K. (2020). *Reuters institute digital news report 2020.* [OCLC: 1203541273].

of the Earth, F. (2017, August 22). *The big ask: How you helped make climate change history* [Friends of the earth]. Retrieved September 9, 2021, from https://friendsoftheearth.uk/climate/big-ask-how-you-helped-make-climate-change-history

Olausson, U. (2009). Global warming—global responsibility? media frames of collective action and scientific certainty [Publisher: SAGE Publications Ltd]. *Public Understanding of Science*, *18*(4), 421–436.

Painter, J. (2013). *Climate change in the media: Reporting risk and uncertainty* [OCLC: ocn841186358]. Published by I.B. Tauris & Co. Ltd in association with the Reuters Institute for the Study of Journalism, University of Oxford.

Palmer, M., Gildea, D., & Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, *31*(1), 71–106.

Panzarasa, P., Opsahl, T., & Carley, K. M. (2009). Patterns and dynamics of users' behavior and interaction: Network analysis of an online community [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.21015]. *Journal of the American Society for Information Science and Technology*, *60*(5), 911–932.

Pearce, W., Brown, B., Nerlich, B., & Koteyko, N. (2015a). Communicating climate change: Conduits, content, and consensus. *WIREs Climate Change*, *6*(6), 613–626.

Pearce, W., Brown, B., Nerlich, B., & Koteyko, N. (2015b). Communicating climate change: Conduits, content, and consensus [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.366]. *WIREs Climate Change*, *6*(6), 613–626.

Punyakanok, V., Roth, D., & Yih, W.-t. (2008). The importance of syntactic parsing and inference in semantic role labeling. *Computational Linguistics*, *34*(2), 257–287.

Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. *arXiv:1908.10084 [cs]*. Retrieved September 10, 2021, from http://arxiv.org/abs/1908.10084

Reynolds, T. W., Bostrom, A., Read, D., & Morgan, M. G. (2010). Now what do people know about global climate change? survey studies of educated laypeople [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1539-6924.2010.01448.x]. *Risk Analysis*, *30*(10), 1520–1538.

Rich, N. (2019). *Losing earth: A recent history* (First edition). MCD/Farrar, Straus; Giroux.

Ruigrok, N., & van Atteveldt, W. (2007). Global angling with a local angle: How u.s., british, and dutch newspapers frame global and local terrorist attacks [Publisher: SAGE Publications]. *Harvard International Journal of Press/Politics*, *12*(1), 68–90.

Ruiu, M. L. (2021). Persistence of scepticism in media reporting on climate change: The case of british newspapers [Publisher: Routledge _eprint: https://doi.org/10.1080/17524032.2020.1775672]. *Environmental Communication*, *15*(1), 12–26.

Samory, M., & Mitra, T. (2018). Conspiracies online: User discussions in a conspiracy community following dramatic events [Number: 1]. *Proceedings of the International AAAI Conference on Web and Social Media*, *12*(1). Retrieved

August 29, 2021, from https://ojs.aaai.org/index.php/ICWSM/article/view/15039

Sawyer, J. S. (1972). Man-made carbon dioxide and the "greenhouse" effect. *Nature*, *239*(5366), 23–26.

Schäfer, M. S., Ivanova, A., & Schmidt, A. (2014). What drives media attention for climate change? explaining issue attention in australian, german and indian print media from 1996 to 2010 [Publisher: SAGE Publications Ltd]. *International Communication Gazette*, *76*(2), 152–176.

Schäfer, M. S., & O'Neill, S. (2017, September 26). Frame analysis in climate change communication. *Oxford research encyclopedia of climate science*. Oxford University Press.

Schäfer, M. S., & Painter, J. (2021). Climate journalism in a changing media ecosystem: Assessing the production of climate change-related news around the world [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.675]. *WIREs Climate Change*, *12*(1), e675.

Schlesinger Jr., A. M. (1986). THE CHALLENGE OF CHANGE. *The New York Times*. Retrieved August 28, 2021, from https://www.nytimes.com/1986/07/27/magazine/the-challenge-of-change.html

Schmidt, A., Ivanova, A., & Schäfer, M. S. (2013). Media attention for climate change around the world: A comparative analysis of newspaper coverage in 27 countries [Publisher: Pergamon]. *Global Environmental Change*, *23*(5), 1233–1248.

Schneider, E. W. (1973). *Course modularization applied: The interface system and its implications for sequence control and data analysis.: (436252004-001)* [Type: dataset]. American Psychological Association.

Schultz, F., Kleinnijenhuis, J., Oegema, D., Utz, S., & van Atteveldt, W. (2012). Strategic framing in the BP crisis: A semantic network analysis of associative frames. *Public Relations Review*, *38*(1), 97–107.

Semetko, H. A., & Valkenburg, P. M. (2000). Framing european politics: A content analysis of press and television news [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1460-2466.2000.tb02843.x]. *Journal of Communication*, *50*(2), 93–109.

Shi, J., Visschers, V. H. M., Siegrist, M., & Arvai, J. (2016). Knowledge as a driver of public perceptions about climate change reassessed [Bandiera_abtest: a Cg_type: Nature Research Journals Number: 8 Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Climate change;Communication;Culture Subject_term_id: climate-change;communication;culture]. *Nature Climate Change*, *6*(8), 759–762.

Shih, C.-W., Chen, M.-Y., Chu, H.-C., & Chen, Y.-M. (2011). Enhancement of domain ontology construction using a crystallizing approach. *Expert Systems with Applications*, *38*(6), 7544–7557.

Singhal, A. (2012, May 16). *Introducing the knowledge graph: Things, not strings* [Google]. Retrieved August 14, 2021, from https://blog.google/products/search/introducing-knowledge-graph-things-not/

Spence, A., & Pidgeon, N. (2010). Framing and communicating climate change: The effects of distance and outcome frame manipulations. *Global Environmental Change*, *20*(4), 656–667.

Sterman, J. D. (2008). ECONOMICS: Risk communication on climate: Mental models and mass balance. *Science*, *322*(5901), 532–533.

Sterman, J. D. (2011). Communicating climate change risks in a skeptical world. *Climatic Change*, *108*(4), 811.

Takahashi, B., & Meisner, M. (2013). Climate change in peruvian newspapers: The role of foreign voices in a context of vulnerability [Publisher: SAGE Publications Ltd]. *Public Understanding of Science*, *22*(4), 427–442.

Tang, S., & Cai, Z. (2010). Tourism domain ontology construction from the unstructured text documents. *9th IEEE International Conference on Cognitive Informatics (ICCI'10)*, 297–301.

Tangherlini, T. R., Shahsavari, S., Shahbazi, B., Ebrahimzadeh, E., & Roychowdhury, V. (2020). An automated pipeline for the discovery of conspiracy and conspiracy theory narrative frameworks: Bridgegate, pizzagate and storytelling on the web (Y.-R. Lin, Ed.). *PLOS ONE*, *15*(6), e0233879.

Tobler, C., Visschers, V. H. M., & Siegrist, M. (2012). Consumers' knowledge about climate change. *Climatic Change*, *114*(2), 189–209.

Van Valin, R. D., Jr. (2004). Generalized Semantic Roles and the Syntax-Semantics Interface [Google-Books-ID: S2OUX4RMK3cC]. In R. Kailuweit (Ed.), *Semantische Rollen*. Gunter Narr Verlag.

Vu, H. T., Liu, Y., & Tran, D. V. (2019). Nationalizing a global phenomenon: A study of how the press in 45 countries and territories portrays climate change. *Global Environmental Change*, *58*, 101942.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks [Bandiera_abtest: a Cg_type: Nature Research Journals Number: 6684 Primary_atype: Research Publisher: Nature Publishing Group]. *Nature*, *393*(6684), 440–442.

Werner, M. D., & Cornelissen, J. P. (2014). Framing the change: Switching and blending frames and their role in instigating institutional change [Publisher: SAGE Publications Ltd]. *Organization Studies*, *35*(10), 1449–1472.

Witte, K. (1994). Fear control and danger control: A test of the extended parallel process model (EPPM) [Publisher: Routledge _eprint: https://doi.org/10.1080/03637759409376328]. *Communication Monographs*, *61*(2), 113–134.

Wu, F., & Weld, D. S. (2010). Open information extraction using wikipedia. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 118–127. Retrieved August 24, 2021, from https://aclanthology.org/P10-1013

Wynne, B. (2006). Public engagement as a means of restoring public trust in science–hitting the notes, but missing the music? *Community Genetics*, *9*(3), 211–220.