Programmable Insight: A Computational Methodology to Explore Online News Use of Frames

by

Loretta Cheeks

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2017 by the
Graduate Supervisory Committee:

Ashraf Gaffar, Chair
Hani Ben Amor
Adam Doupe
Nancy Cooke
Dara Wald

ARIZONA STATE UNIVERSITY

May 2017

ABSTRACT

The Internet is a major source of online news content. Online news is a form of large-scale narrative text with rich, complex contents that embed deep meanings (facts, strategic communication frames, and biases) for shaping and transitioning standards, values, attitudes, and beliefs of the masses. Currently, this body of narrative text remains untapped due—in large part—to human limitations. The human ability to comprehend rich text and extract hidden meanings is far superior to known computational algorithms but remains unscalable. In this research, computational treatment is given to online news framing for exposing a deeper level of expressivity coined “double subjectivity” as characterized by its cumulative amplification effects. A visual language is offered for extracting spatial and temporal dynamics of double subjectivity that may give insight into social influence about critical issues, such as environmental, economic, or political discourse. This research offers benefits of 1) scalability for processing hidden meanings in big data and 2) visibility of the entire network dynamics over time and space to give users insight into the current status and future trends of mass communication.

DEDICATION

This research is dedicated to the loving support and encouragement of my father and mother, Albert and Joyce Henderson, who taught me the value of faith, perseverance, and humility. My mother-in-law, Mary P. Cheeks, gave consistent encouragement and love on this Ph.D. journey. Their wisdom is forever etched in my heart and mind. This work is also dedicated to my loving husband, Michael L. Cheeks; my sons, Jordan and Ashton Cheeks; and my daughter-in-law, LaToya, who have been a great support. I also dedicate this work to my granddaughter, Harmony Ira Cheeks, who has inspired me to keep a clear lens in which to view the universe and continue to discover the bounty of life. To my mentors, Dr. Albert McHenry and Dr. Dara M. Wald, I dedicate this research. They were a tremendous help navigating the Ph.D. process.

ACKNOWLEDGMENTS

I thank my God and my Savior for giving me the vision, purpose, and steadfastness to finish my course with joy and a renewed sense of awakening. Special thanks to the GEM Consortium and Adobe Foundation for being a rainbow in my cloud. I am eternally grateful to Dr. Ashraf Gaffar for being an excellent advisor, charting a path for my completion, and seeing the value of my research. I wish to thank my amazing collaborators, Dr. Tracy L. Stepien, Dr. Derdei Bichara, and Dr. Dara M. Wald for investing countless hours to this research. I would like to thank Dr. Nancy Cooke, Dr. Heni Ben Amor, Dr. Adam Doupe′, and Dr. Dara M. Wald for serving on my committee and for their valuable feedback. I thank my good friends Ardeana “Dina” Weber and Cheryl Lynn Jones for always reminding me that I can do this and for being so kind and loving. Dina, your encouragement and words of wisdom will forever be treasured. Cheryl, your words, “finish strong” and “fear and faith cannot occupy the same space,” lifted me when I grew faint. I am so grateful to my amazing and supportive family for demonstrating comeback power and resilience in the face of adversity. I would also like to express special gratitude to my cousin Michael Pittman for his support and prayers. Special thanks to the Center for Policy and Informatics, who participated in this research. The timelines of your contributions were noteworthy. I am extremely grateful to Charles Moore and Barbara Moore for their editing assistance.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

I. INTRODUCTION

Approximately 80% of all data today exist in digital form as unstructured text (e.g., news, e-mails, social media feeds, contracts, memos, clinical notes, and legal briefs) [52]. Much of this unstructured text exists on the Internet or is transported using the Internet. The Internet is the premier platform for the proliferation of online news and has been compared to a huge social and psychological laboratory [40]. It has enabled a highly global and connected world, which is greatly transforming computational approaches for exploring online social interactions and, therefore, society. News sources disseminate online news that operates within a complex and dynamic system via the Internet. This system comprises informational flows, interactions, and structural variations, which lead to asymmetries and have a tendency to influence society. Online news is a form of narrative, unstructured text that embeds facts and strategic communication frames for shaping and transitioning group standards, values, attitudes, and beliefs about critical issues.

In accordance with D. P. Cartwright, ″The formation and transmittal of group standards, values, attitudes, and skills are accomplished largely by means of communication″ [53]. Consequentially, efforts to understand social interaction, cooperation, conflict, and influence require the study of text. Given the widespread use of unstructured text for individual and mass communication, such as email, online news, and social media feeds, it is particularly important to understand how social influence and information about complex, critical issues spread through online content over time.

Although communication research on the use of frames and public attitudes in traditional news venues has been widely explored, the identification and analysis of frames in unstructured online news text has received minimal attention. This gap in research exposes potential opportunities for advancing computational tools for synthesizing knowledge sources and for deep learning to improve machine intelligence that makes room for the automatic emergence of knowledge. This is an increasingly important area of research, given the vast amounts of daily production of online news and the challenge of humans to make sense of these data for making decisions. With existing computational tools, it is unknown whether sheer volume, access, and availability of data translate added value to news sources and to readers for making informed decisions and gaining a clear understanding of critical issues. Furthermore, with existing computational tools, it continues to be a challenge to progress toward learning the latent meanings of more advanced forms of narrative text, such as poetry, metaphors, jokes, etc. The ability to unlock hidden structures and latent meanings in narrative unstructured text and to explore influence and migration paths of text are important areas of research because text is a fundamental device of communication and human interaction for expressing real-world issues.

The embedded facts that exist in narrative text represent absolute, relevant truths characterized by their rigid nature. These absolute truths are shared among news sources. However, purely factual articles would be hard to write and read. The additions of writing style, opinions, related context, and nonfactual comments and arguments often make text more subjective but easier to read. The exercise of news framing—and the amplification of subjectivity that it conveys—is where multidimensional aspects of the content emerge

and where expressivity happens for transformational and maximal influence. Framing is the use of strategic devices for presenting prominent aspects and perspectives about an issue using certain words as well as stereotyped images and sentences for the purpose of conveying hidden meanings about an issue [30]. News sources play a critical role in the activation of frames about issues, and the amplification of subjectivity in online news drives a wedge between evidence and beliefs [36]. In addition, the combination of facts, frames, and amplified subjectivity represents instruments that offer a space of possibilities for deeper exploration of the cognitive aspects of narrative text. Natural language processing (NPL) has been used to make significant progress, using artificial intelligence (AI) to parse narrative text at a syntactic level. However, less success has been made with computational treatment when it comes to semantic levels. Semantic parsers work well only at a phrase level, but semantically parsing several sentences together is challenging. Some approaches can be used to look for keywords to determine the sentiment of a text, to provide a summary, or to provided language-to-language translation, but they remain limited and are not embedded with the human-level ability to understand the deep meanings of a complete text or of a related set of articles. I define this human-level understanding as an ″insight,″ which is ″the ability to understand the latent meanings within an article.″ Insight can also be generalized to apply to a group of related articles and refers to ″the ability to see the relationship between a group of articles as well as how the contents of later articles were affected or influenced by those in earlier articles of the same group.″ The need to understand this insight becomes particularly

important with online news, as the number of related articles tends to be generally larger than in printed media due to the far reach and the larger readership of the Web. This understanding process is typically highly cognitive and is best done by humans. However, with the large number of online news, the need for software tools becomes essential for gaining programmable insight. I define programmable insight as ″the ability to apply computational methods and, hence, software tools to leverage human cognitive ability to gain insight into narrative text.″

In attempts to identify potential pathways to address the complexities and dynamics that exist in online news text, frames are used as instruments for experimenting and learning about the production of language for gaining programmable insight that may lead to new instruments for exploring cognitive computation. I recognize the need for an integration of multiple models and perspectives that advance what is known about hidden meanings and latent structures in text. Modeling narrative text allows software tools to provide help in automating human efforts using computational leverage. However, providing a deep model that represents hidden meaning in text is relatively hard. Several NLP methods in AI use word matching and probability functions to determine the correctness and possible meaning of words with the help of language corpora containing trillions of words. This ″brute force″ approach could be computationally heavy, and it does not provide an insight beyond the phrase level. An acknowledgment of the multidimensional aspects—degrees of freedom—of the narrative text in news calls for

computational research to fill knowledge gaps and to transform the current understanding of alternative pathways to learning that leverage cognitive science research.

Leveraging the scholarly work of mathematicians, the fathers of the digital computer developed the initial logical machine using a rather formal grammar structure. The first level of grammatical expressivity the computer machine could understand is propositional logic—that is, knowledge of facts or truths. Considering the ″form″ of the English language (i.e., subject, verb, object), the logical machine could make valid arguments using as few as two premises and simple symbols, thereby drawing a logical conclusion. Propositional logic became the basic component of symbolic logic. It became the enabler for the logical machine to reason using simple English grammatical structure. This breakthrough led to machines being able to reason as either true (1) or false (0), using simple symbols to express ″and,″ ″or,″ ″not,″ ″implies/then,″ and ″if and only if.″

By switching out fixed sentence structure for symbols, as was done in mathematics for centuries, the logical machine could express a primitive level of truth. The computer could express with ″0″ or ″1″ with this simple syntactical machine code. Although this advanced in the logical machine was very impressive in its day, this level of formalism with its strict grammatical structure falls short of reaching the goal of AI, as it gave no insight into the topic proposed in the sentence structure.

Propositional logic usage in the logical machine had no properties nor a sense of relationships, all of which were embedded in the internal structure of the text. Rather, its sole focus was on the sentence structure as an object for expressing what logicians for centuries thought to be real. Real objects were point masses and forces; as such,

properties leading to an understanding about the relationships and internal structures of text were dismissed as subjective. Discounting relational properties as the inferior status of ″secondary″ properties resulted in excessive adherence to the theoretical approach of attaining machine intelligence. However, researchers soon came to realize that if there is any future for theory, it must be applied. This meant that the internal structure in text must delve into deeper levels of expressivity where hidden meanings and relationships could be explored for advancing machine intelligence. Since this time, as shown in Fig. 1, numerous levels of machine logic have been established to allow for the variety of properties to reason about relationships, individuals, identities, and quantifiers—thereby, giving way to make more elaborate assertions, inferences, and semantic expressions of phrases. Fig. 1 shows the progressive climb up the stack from the formal space where computer machines focused on propositional logic, formal grammar, and, thereafter, phrases.

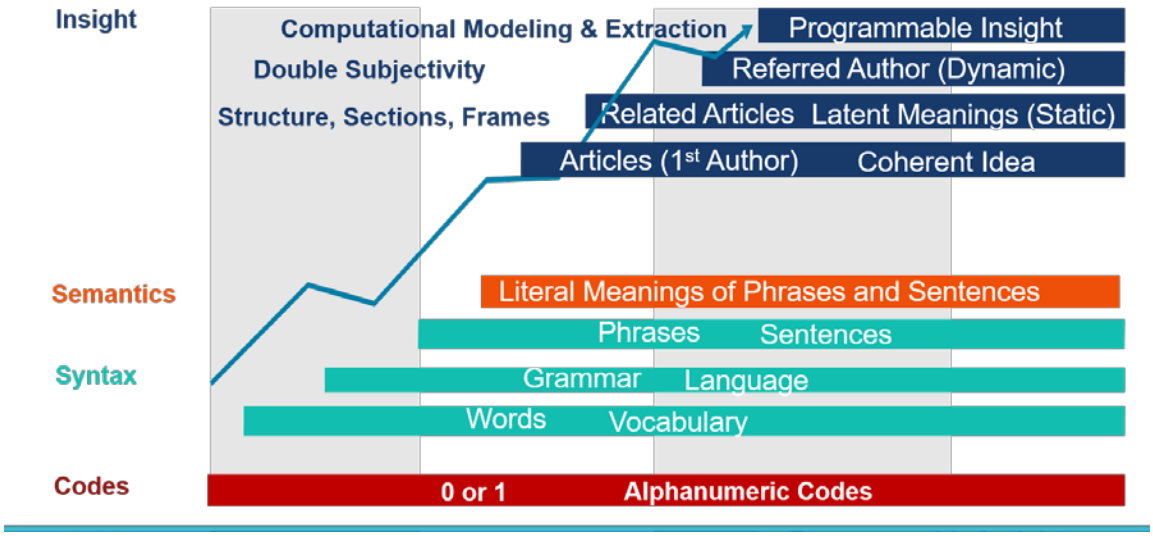


Fig. 1. A view of the end-to-end bridge to cognitive computation model. Here, the stack, from bottom to top, represents the evolution of computational processing from grammar to insight, as shown on the right, and the value or outcome that results as shown on the left. For instance, computer processing of 0 and 1 bits were the catalyst for primitive storage, transmission, and processing logic.

With the pendulum swinging from machine-to-human logical expression—where the programmer of the computer machine had to learn the psychology of the machine—to human-to-machine logic for expressing rational human thoughts, higher-level programming languages were developed, such as Fortran, Prolog, and Lisp. With the foundations for reasoning and with higher levels of expressivity at the researcher′s grasp, attention turned to another field of research, AI.

Here, at the phrases layer (i.e., at AI) is where much of the computer science research today continues to focus. This has led to great breakthroughs for learning AI, NPL, and fuzzy logic; however, this approach shed very little insight into cognitive computation.

This research advances beyond the phrases layer and bridges the computational and cognitive worlds using the instrument of frames that exist in narrative text for exploring programmable insight that may lead to deeper learning for machine intelligence. In exposing the discovery of new knowledge about text, transformations (or shifts) in which dynamic forces reach a critical point make clear the new structures, pathways, and influence patterns to be accepted or new questions or alternatives in the scope of an issue to be revised. The ability to seek out novel, uncharted relationships that are implicit from a body of unstructured narrative text—yet not explicitly stated within that text—has obvious scientific value. This process is called "novel knowledge discovery." It enables the use of the current state of knowledge for an emergence of possible new relationships and pathways that have yet to be studied.

This dissertation includes an exploration of programmable insights into online news sources' use of news frames within articles—a form of narrative text—to expose a deep level of expressivity—what is referred to in this paper as "double subjectivity," a term introduced here. The aim is to demonstrate news framing as a step toward deep learning about the production of language that may inform scholars about intelligent behavior in machines.

A. Structured Expert Support Models

Fully structured data render itself easily to computation. A good example is the sales receipt generated by point-of-sale (POS) software in most retail stores, showing a list of the code, text, sale price for items as well as the subtotal, total, date, and time of each receipt. Terabytes of such data follow a well-known model, they would be easy to

process with software, and they provide a solid foundation of business intelligence and other advanced insight. Narrative text, however, is highly unstructured and could represent a nightmare to computational algorithms used with the aim of extracting deep knowledge buried within. One of the main reasons for the complications is the lack of a representative model that these data follow. Although humans can understand and even enjoy the contents of such text, the lack of concrete models prohibits the computational treatment of such text. In the approach put forth in this research, the framing theory is used—as will be shown later—to build representative models of narrative text. Although these models need to be built by hand using human experts (i.e., expert support), they are done in a structured way to normalize human cognitive ability into nominal structure (hence, the qualifier Structured Expert Support [SES]), as shown in Fig. 2. Humans will manually extract certain keywords from a text and build data structures (vectors) representing hidden meanings within the text. These vectors will be a strong representation of the hidden frames in a text, and they will be highly structured and, hence, readily computational. As will be discussed, the SES relies heavily on the theory of framing as the modeling concept and as the bridging between cognitive contents and computational capability of narrative text.

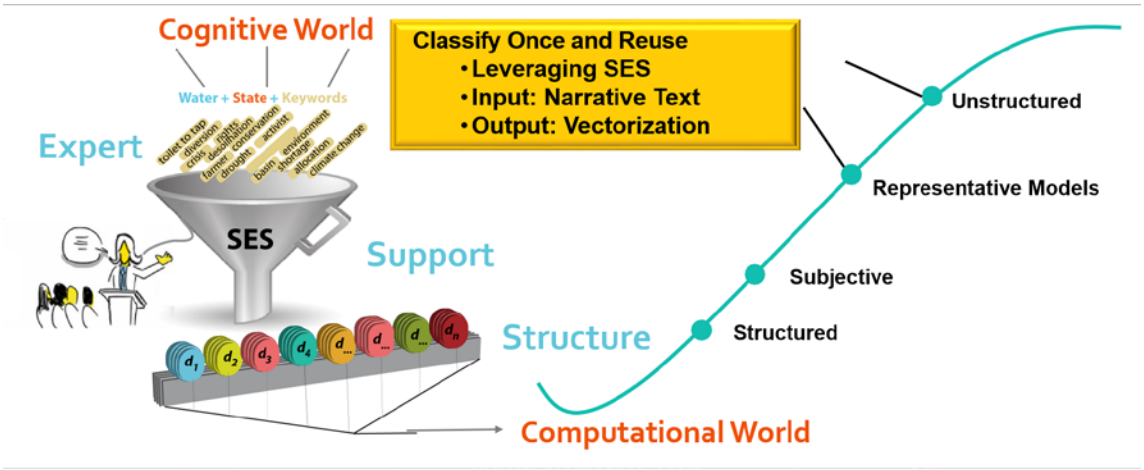


Fig. 2. Depiction of SES significance in the development of a representative model in support of the discovery of frames that exist in unstructured narrative text.

In exposing the discovery of new knowledge about text, transformations (or shifts) in which dynamic forces reach a critical point make clear the new structures, migration paths, and influence patterns. The ability to seek out emergent structures that are implicit from a body of unstructured narrative text—yet not explicitly stated within that text—has obvious scientific value. This process is called "novel knowledge discovery." It enables the use of the current state of knowledge for an emergence of possible new relationships and pathways that have yet to be studied.

This dissertation includes a model for exploring frames as instruments with online news articles—a form of narrative text—to expose a deep level of expressivity—what is referred to in this paper as double subjectivity. The double subjectivity model [50] is presented for gaining programmable insight into the hidden structures in text and understanding influence patterns. The double subjectivity method expands Axelrod′s [7]

and Klienberg′s [28] research on cultural dynamics and social influence for constructing an issues network representation and designing a visual language for exploring spatiotemporal dynamics. The issues network is exemplar of a complex dynamic system whereby migration paths may be visualized. The migration path is defined as the movement of frame choices in articles across a network to establish a new bridge for information exchange. The ways that the issues network grows (gain node connections) and shrinks (lose node connections) are discussed in a subsequent section in this manuscript: "Double Subjectivity Social Influence Model." The aim is to demonstrate news framing as a step toward deep learning about the production of language that may inform scholars about pathways that offer new approaches that advance cognitive computation for improved intelligent behavior in machines.

To the best of my knowledge, in computer science scholarly research, examining double subjectivity using online news and framing effects for gaining programmable insight has not been addressed previously for the class of problems discussed in this paper.

B. Conceptual Framework

In the last decade, advances in fields such as AI, text summarization, and NPL have enabled gains in understanding text syntactic structure and limited gains in text semantics by offering algorithmic and theory-based methodologies. However, these fields have struggled in efforts to unlock hidden meanings and abstractions found in unstructured text such as jokes, poems, or art that are readily expressed through cognitive structures. This is due in large part to barriers that exist when using machine learning algorithms that are still limited in their understanding of complex unstructured text where factors such as

semantic context, culture, and complex relationships must be considered for making sense of the narrative text.

Humans still have the unique and superior ability to get beyond the syntax of data in order to make meaning. However, a great deal of cognitive science research has focused on cognitive structures called “frames” (sometimes “schemas”) that exist in the recesses of the brain [51]. The framing of online news articles by news sources consequently has the effect of evoking a word and phrase in the unconscious cognitive structure in the mind of the reader. These structures are built over time through repetition. The repetitive framing of critical issues will strengthen the circuits about the issue in a reader′s brain. When online news sources repeat frames often, they become ″normalized″ unconscious activation points in the brains of readers. This behavior by online news sources is equivalent to placing ideas in a container, referred here as an article. These ideas are sent through communication channels that have tremendous influence on the perception and real value of knowledge. When neighboring news sources take the ideas out of the container by way of adoption or copying, the cumulative effect is double subjectivity. A property of double subjectivity is spatiotemporal dynamics, which has the capacity to display a larger context of influence. As a number of related articles grows, the cumulative effect of subjectivity could grow and take on certain direction or shape, influenced by the article-to-article changes or the amplification or suppression of frames. This dynamic behavior may be best shown through a network visualization. A network visualization is a mechanism that has been used to leverage people′s ability to acquire

knowledge for revealing hidden meanings, social influence, and migration paths of information and interactions.

The automation of unstructured text can be generally categorized under two main areas: computational data and cognitive data. Computational data—represented by mathematical formulas, propositional logic, and formal (computer) languages—have little room for subjective interpretation and are typically concise and predictable. For instance, a correctly developed source code can compile and run on a computing machine and will always produce the same output for a specific input. The combination of a formal set of instructions resulting from the successful compilation and run of source code can transition a computer through a set of predefined states, depending exclusively on the compiled code and input data. This leaves no room for interpretation or subjectivity. In other words, the meaning and interpretations must be fully represented in the source code. A good programming language must, therefore, have sufficient expressiveness to fully represent a feasible solution in a specific problem space. If a source code arrives at an undefined state, it will simply freeze or crash; it is unable to make an independent decision. Some areas of computer science such as fuzzy logic and AI attempt to address these inherent limitations. However, these approaches are not the focus of the work conveyed here.

Although the computational language of source code is a good medium for communicating with computers, natural languages that embed cognitive structures are the mediums of communication among humans, and natural languages are informal and ambiguous in comparison to the computational language of computer code. Cognitive

data can be further divided into two subcategories, depending on the source of origination: direct and indirect.

Direct cognitive data are the result of a person′s encounter with an event and the perception, in part, of its details firsthand. Human perceptions of events lead to personal interpretations of events. Interpretation depends greatly on the viewer′s perception, as different people can arrive at different interpretations of the same event. Perception is often associated with the viewer′s state of mind, prejudice, background, and beliefs. Two persons witnessing the same event or being involved in the same situation could assign different interpretations and end with separate conclusions. Although the external situation is the same (the perception of which is uniquely defined in this paper as the absolute baseline), it could be internalized differently by different people. Such a case is a first-level subjectivity. Even when different people are asked to report an incident or a situation (the same absolute baseline), different accounts are often the result, indicating first-level subjectivity. Many factors further affect this subjectivity:

- Subjectivity: Humans have different expressive abilities and styles in reporting events using natural languages, leading to increased subjectivity.
- Ambiguity: The ambiguity of natural languages allows for different interpretations of spoken or written sentences.
- Complexity: Complex, extended situations, such as movie showings, are harder to perceive and interpret compared to simple or atomic events like the occurrence of a rocket launch.

- Expressiveness: Unlike a computational language, a natural language does not fully express the contents in its syntax. Natural languages report about something, rather than fully describe an action, artifact, or event in a specific way.
- Context: Natural languages are highly context dependent. The same sentence could have different meanings in different contexts.
- Modality: Natural languages are often combined with other modalities like paraverbal communication, body language, and face impressions, which can influence the meanings or interpretation of spoken words.

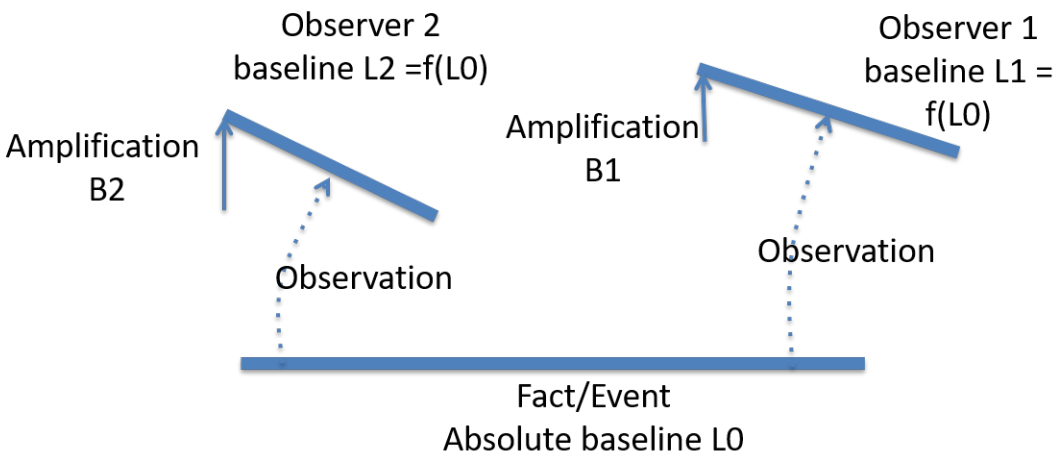
The aforementioned factors explain why—even for different people directly witnessing the same situation—substantial subjectivity can be present in reporting or discussing an event that was mutually witnessed.

People often try to reduce first-level subjectivity. For example, television reporters try to abide by certain rules to maximize neutrality and leave it up to the viewers to build their own opinions freely. The scientific community tries to show the stated facts through rigorous research and scientific experiments in attempts to eliminate or reduce subjectivity. However, subjectivity is not so simple to discard. Consider newspaper and television reporters who are known for their opinions.

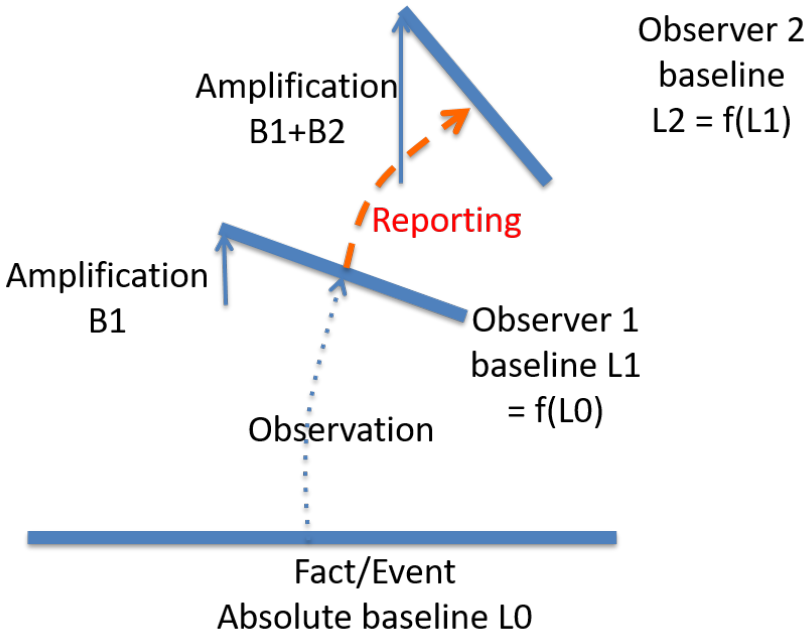
The Internet and the worldwide web have led to an explosion of indirect cognitive data although events can still be directly perceived and subjectively interpreted. Because of technology advancements, it is possible to receive a much larger number of events indirectly. Because these events are not directly witnessed, an absolute baseline is not

available. Because knowledge of an event is often based on the written or spoken reports of others, the baseline will be relative to the reporter′s subjectivity or simply "float."

With the acceptance of indirect reports and their interpretation, the receiver adds his or her own subjectivity. Again, no longer is the absolute baseline available. It is replaced with a reporter-relative baseline or a floating baseline. If the reporter′s subjectivity is strongly framed and influenced towards certain beliefs, the receiver of the account already starts with a baseline that is completely different from the absolute baseline. This new baseline could strongly affect the receiver′s beliefs. Double subjectivity is the term introduced in this paper to describe the secondary subjectivity that results after receiving indirect informal data. Fig. 3 illustrates the cumulative influence effects of double subjectivity.



(a)



(b)

Fig. 3. Two representative illustrations of the double subjectivity framework. (a) There is an event (Observation) about an issue that two observers see: Observer 1 and Observer 2. Although it is the same event, their perceptions are different. Their amplification may be different, but independent. (b) This graph illustrates the introduction of an indirect account of an event, whereby the reporter does not receive the information firsthand. Rather, Observer 2 adopts the report of Observer 1, which sets into motion the cumulative effect of double subjectivity.

The illustrated formula depicts a generic function for double subjectivity in online news. Here, the assumption is that there is an observer-dependent function $f(x)$, such that for first-level subjectivity, it is expressed as,

$$L_1 = f_1(L_0), \text{ and} \quad (1)$$
$$L_2 = f_2(L_0)$$

whereas, double subjectivity is expressed as,

$$L_1 = f_1(L_0), \text{ and} \quad (2)$$
$$L_2 = f_2(L_1)$$

The amplification intensity function f depends on the person observing, then reporting a situation or event.

An analogy drawn from photocopying illustrates double subjectivity in that, ideally, a photograph copy should be a close replica of the copied photograph. Yet, in actuality, the resulting photograph is the copied photo with its amplification plus the distortion amplification gained from the copying process or machine.

Online news is exemplary of double subjectivity in that its assemblage consists of subjective properties such as context, interpretation, multimodal interaction, background, compensation, and assumptions. These properties are critical building blocks that characterize narrative text for its 1) signification (i.e., the semantic content signifying an association or concept given a word, sentence, or phrase) and 2) significance (i.e., the relevance, rank, importance, or capacity to make a difference). Unfortunately, substantial gaps remain in advancing what is known about the production of language in narrative text where double subjectivity exists. Therefore, exploring deeper levels of expressivity with an emphasis on double subjectivity may inform a new direction in research.

Discovering double subjectivity in narrative text poses a difficult problem, as it provides insight and meaning to open-ended statements and indirect data. More complex structures within the narrative text provide an expansive landscape for advancing effective tools and crafting applications to automate many language-related tasks—for example, document summarization, automated text generation, and many others. To show the gains in logic and reasoning, Fig. 4 depicts advances made using current tools and new frontiers that double subjectivity may shed light on, such as poems, metaphors,

idioms, jokes, art, and storytelling. These advances are to be achieved by exploring levels of narrative text.

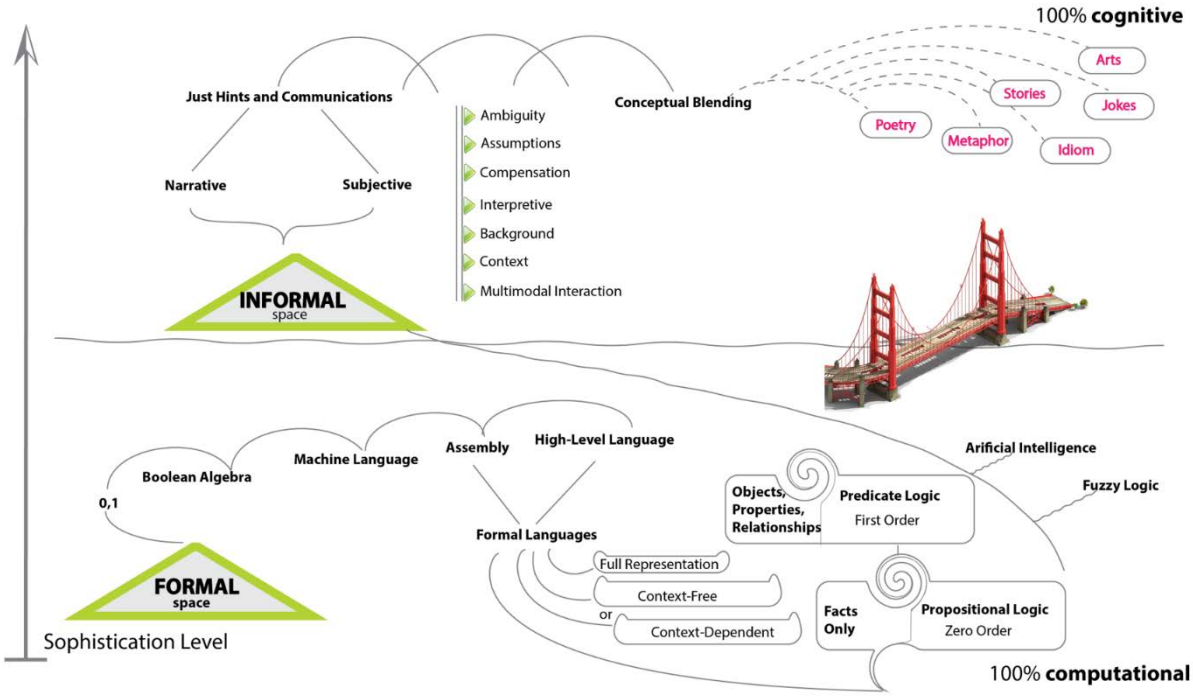


Fig. 4. Depiction of the bridge from the computational world to the cognitive world for exploring first-level subjectivity and double subjective potentials.

1) Problem Statement: This dissertation gives computational treatment to online news sources′ use of news frames for gaining programmable insights. Here, algorithms, methodologies, a network, models, and a visual language are offered that build on principles of machine learning, graph theory, social influence, and cultural dynamics for unveiling hidden meanings in narrative text and instruments for deep learning.

2) Discovering Frames in Online News: To date, there is no formal definition that the computer can understand for discovering distinct news frames. This research fills this gap by formulating a formal definition and algorithmic approach for discovering

news frame in online news text. Discovering news frames is an important area of research in that it is an alternative way to unveil latent meanings in text. In the problem below, the formal definition of framing is presented.

Problem: Let U_t be a universe of online news documents at time t. Let $\{d_i\}, 1 \leq i \leq n$, denote a finite subset of documents from the universe of documents U_t, where n is the total number of documents being considered. Let $S_i = \{P_i, C_i\}$ be the set of document properties, where P_i is the source that produced the article and that propagates a central organizing idea (i.e., framing) and C_i is the set of terms (i.e., features or content descriptors) from the news article where common words have been removed—all of which supply a context and suggest what the issue is through the use of selection, emphasis, exclusion, cues, and elaboration [54]–[55]. For the issue of water insecurity, I use standard frames dominantly used in U.S. news coverage [56]: human interest, conflict, economic, managerial, and science. Let $f_1, f_2, f_3, f_4,$ and f_5 correspond to each of these frames, where each f_i is associated with a pair $\{x_i, y_i\}$; $\boldsymbol{x_i}$ is the feature vector, and $\boldsymbol{y_i}$ is the vector with the corresponding weights. Discover five dominate news frame signatures.

3) Frame Mapping Approach:Traditionally, framing effects within media communication have been examined using content analysis, which was developed specifically to aid in the interpretation of social discourse or text for communication research [49], [57], [58]. Content analysis involves methodical evaluation and the systematic assignment of communication content to categories per rules and the analysis of relationships involving those categories using statistical methods [59], [60]. Current

approaches to content analysis, however, require scholars and researchers to thoroughly examine documents in search of patterns for the categorization of frames that exist in the news text. This approach to text analysis is dependent on humans and limits the applicability of the analysis of large-scale unstructured text. Thus, a more effective tool for unlocking latent meanings found in unstructured text could provide a mechanism for delivering knowledge in online news text and thereby enhance an understanding of online behaviors and lead to deeper levels of expressivity. Here is a new approach to validate the discovery of distinct frame patterns for extracting latent meanings that exist in online news articles.

Problem: Let U_t be a universe of online news documents at time t. Let $\{d_i\}, 1 \leq i \leq n$, denote a finite subset of documents from the universe of documents U_t, where n is the total number of documents being considered. Map $f_1, f_2, f_3, f_4,$ and f_5 that correspond to human interest, conflict, economic, managerial, and science frames to each d_i for determining d_i classification and validation of d_i against the f_i frame signature that functions as the ground truth.

4) Online News Issues Network: Model a news frame issues network as a graph $G = ({S_g}^t, {A_h}^t, {E_r}^t)$ consisting of 1) S_g, which denotes a set of source nodes containing $g = \{s_1, s_2, \cdots, s_g\}$ elements and 2) A_h, which is a set of article nodes containing $h = \{a_1, a_2, \cdots, a_h\}$ elements that operate within the time window t. In general, t indicates one discrete time step of a reporting period by S_g. Let $S_g = \{C_i, F_i\}$ be the set of source properties, where 1) C_i is the set of terms (i.e., subissue keywords or content descriptors) and 2) F_i is the dominant frame choice (identity) that the news source

uses. Let R be the number of relations. Each relation has a corresponding set of edge connections, ${E_r}^t$, directed/undirected as edge elements. The issues network, subscript $r = 3$, corresponds to the total number of relations. The edge set represents the communication channels between node pairs. A_h are children of S_g; therefore, they may inherit the properties of the parent news source, as shown in Fig. 1.

The relations (or rules) for E_r edge connections are as follows:

$\boldsymbol{R_1}$. A & A (articles to articles). A nondirectional edge connection is constructed for $E_1$when $A_h(a) = \{a: (a_1, a_2, \cdots, a_h) \in E_1\}$ is the set of neighbors of $\boldsymbol{a_h}$ such that $\boldsymbol{a_h}$ represents articles with the similar subissue.

$\boldsymbol{R_2}$. A & S (articles produced by the same source). A directional edge connection is constructed for $E_2$when $S_g \in E_2 \rightarrow A_h(a) = \{a: (a_1, a_2, \cdots, a_h) \in E_2\}$ is the source (or producer) of the set of neighbors of $\boldsymbol{a_h}$ articles, where $\boldsymbol{a_h}$ represents similar article subissues.

$\boldsymbol{R_3}$. S & F (sources with subissue to subissue with the same frame choice). A nondirectional edge connection is constructed for $E_3$when $S_g(s) = \{s: (s_1, s_2, \cdots, s_g) \in E_3\}$ is the set of neighbors of $\boldsymbol{s_g}$ such that $\boldsymbol{s_g}$ shares the same dominant frame choice.

Problem: Given $G = ({S_g}^t, {A_h}^t, {E_r}^t)$ about I issue, construct a graph for determining how frames are produced in online news media over time.

5) Double Subjectivity Social Influence Model: The double subjectivity social influence model integrates cultural dynamics extensions [7], [28] with the cusp catastrophe model (CCM) [38], [40] for modeling double subjectivity in narrative text.

The CCM is useful for describing nonlinear relationships such as those found in narrative text. The CCM allows for understanding the forces at work when quantifying amplification and the migration path of amplification formation. Online news sources tend to produce articles about an issue I of interest to society, such as water insecurity, the state of the economy, or affordability of health care. Additionally, news sources may receive signals β from other news sources on the importance of the issue, based on observed increase in the number of articles other neighboring news sources produce. They receive facts f about an issue. Each news source, then, selects an amplification strategy α using as the basis of its calculation the amplification intensity. More specifically, each news source S_g with interest in I issue at time t generates news articles

$$I(\mathrm{g}) = \alpha_g + \beta_g(f) \tag{3}$$

where α_g and β_g are the coefficients of the news source and f represents the facts received. Time evolves in discrete steps t = 0, 1, 2,⋯, and $S_g(t)$ denotes the mass count on S_g at time t.

This research situates the quantification of amplification intensity in terms of a mathematical catastrophe model [38], cultural dynamics [7] and social influence models [28]. This measure is important for determining the mass, meaning size of the nodes in the network, growth, and shrinkage through adoption. Traditionally, researchers have incorporated the empirically well-supported observation of cultural dynamics and social influence through the lens of pluralism, which follows a power law distribution. These models can be best characterized as the rich get richer phenomena, as the Matthew Effect, or homophily, i.e. the tendency of birds of a feather to flock together. These models that

have been used to examine social influence dynamics imply that monoculture is unavoidable, unless a subset of the mass is perfectly cut from outside influences. I argue that there may be an alternative possibility space that can be beneficial for understanding framing effects within the cultural dynamics and social influence phenomena, which consider underpinning exogenous forces that affect the mass and the diversity in selection and influence that may happen through interactions in a network. Particularly, when considering the vast amounts of online news and the interconnected news sources, these factors may allow new pathways to be explored for learning about the production of language. Thus, the integration of the CCM with the cultural dynamics model contains the properties needed for exploring greater degrees of freedom where a combination of forces is at work. Here, a model is presented that bridges the computational and cognitive worlds that align with the understanding of human behavior, particularly when the framing of an issue is repeated—this coupled with research about the behavior of the users of Internet, where strong correlations exist between the expressed opinions and copying effects (e.g., fake news) that shift reader′s attitudes about critical issues. The model explores three variables for examining gradual changes that affect frame amplification intensity. They are amplification for showing attitudes or beliefs shifts, justice or a sense of fairness, and frame identity.

Problem: Given $\{S_g\}$ interactions within graph G_t over time, determine the migration path and the nonlinear, causal linkages about $I(\mathrm{g})$.

6) Visual Language Design for Reading the Model: I designed a visualization of the

dynamic graph under observation for showing the cumulative effects of double subjectivity that give insight into the issue′s spatial and temporal dynamics that are relevant to social influence. The visualization offers a human-suited representation of the network structure, the rhythmic timing, the growth and shrinkages of the node masses, and the interplay of forces that may give insights into the extent of news sources′ involvement in shaping attitudes, values, and beliefs about critical issues over time.

C. Research Questions

A content analysis supported the intent for the current research, which was to assess the effectiveness of giving computational treatment to discover the usage of dominant frames by online news sources and the social influence dynamics that occur through the propagation of vast amounts of news over time. Evidence supports the dominant influences of managerial frames, which give insight into the prominence of elitists in shaping attitudes, beliefs, and values about critical issues. Although this evidence from the narratives of managers or authoritative figures is important, the research inquiry is an examination of the use of the five standard communication frames for giving insights into the spatiotemporal dynamics. This research inquiry includes an overarching question: ″How is the issue of water insecurity in the Southwest region of the United States being framed in online news media over time?″ This research inquiry will give insight into the following questions:

1. RQ1: How do online news frames emerge?
    a. Which dominant frames emerge from the unstructured text?

b. How is framing used as a mechanism for conveying latent meanings and hidden structures in online narrative text?

c. How is framing being produced?

2. RQ2: What is the migration path of online news frames as they evolve over time?

    a. What is the effect of double subjectivity?

    b. What are the sizes of the masses of articles by frame type?

    c. Do the masses grow or shrink over time?

3. RQ3: What are the conditions observed for determining shift in beliefs?

    a. What is the measure of amplification intensity embedded in the text?

    b. What are the implications of frequent usage of frame strategies?

    c. What are the forces that cause frame amplification shifts to occur?

The objectives of this research inquiry are threefold: 1) to contribute to the development of deeper levels of learning through the unveiling of a new conceptual framework coined “double subjectivity” by which online news sources gain the cumulative value of amplification when they embrace the indirect account of an issue (or event) from another news source for the generation of news articles; 2) to explore the internal structure of narrative text through the discovery and examination of news frames for scaling up to large datasets; and 3) to explore the effects of online news sources′ use of framing in propagating news articles for shaping beliefs, attitudes, and social norms about critical socioenvironmental issues. The linkages between news framing effects for influencing society have been studied extensively in communication and social science research. Only recently have news frames been the subjects of research in computer

science; hence, much work remains to improve the understanding of the use of news frames as devices for exploring the production of language in text.

D. Contributions

This dissertation claims the following original contributions:

1. Chapter 3 includes the definition of the first computational formalization of news frames. Here, this research leverages machine learning and mathematics to test a new approach for automatically discovering distinct frame patterns in online news using a process that advances semisupervised machine learning for document clustering and classification. This approach for automatic discovery of news frames advances learning for document classification.
2. In Chapter 4, a new conceptual framework, double subjectivity, is presented and is expressed through a news issues network. The principal property of double subjectivity is its cumulative effect for the spread of news narratives.
3. In Chapter 5, this dissertation details the double subjectivity social influence model for expressing the multidimensional aspects of news framing—and the amplification that it conveys—for transformational and maximal influence. The definition of the model, as expressed in this chapter, involves the computational formulation of amplification intensity as a function in operation in the narrative text.
4. In Chapter 6, this research offers the design of a visual language for exploring the news issues network and model dynamics that provide

programmable insights for unveiling hidden meanings in narrative text using as its principal instrument news frames.

E. Broader Impact of This Work

Media attention and framing can influence public and political concern about a specific topic. In order to observe the potential pathway of cognitive, informal, natural language, I took as a subject water policy, a subtopic in the environment domain. The framing of an issue has an impact on the ultimate course of the issue. The ways that water policy and decisions are framed affect water rights allocations, policy decisions, human consumption, emerging technologies, farming techniques, and agricultural outcomes [29].

Water is a fundamental resource affecting all aspects of life on earth. Water is used for human consumption, industrial processes, production of food, sanitation, as well as other uses. The issue of water insecurity in the U.S. Southwest region is particularly important in that seven states (Arizona, California, Colorado, Nevada, New Mexico, Utah, and Wyoming) rely primarily on fresh groundwater flows originating from the Colorado River. The Colorado River was a sufficient source of water for decades; however, due to a decade of drought, this once plentiful source of water cannot meet the demand to sustain life as it exists in this region. This situation contributes to water insecurity throughout a large region of the United States.

Water insecurity in states that source their water from the Colorado River exemplifies an issue that is appropriate for study through the lens of double subjectivity and that can be observed via the framing strategies used by online news sources. The topic of water insecurity is ideal because this issue is context specific, complex, and characterized by multiple interacting forces. Newspapers in the Southwest regularly

produce in-depth articles about drought, water, and climate change that are published online. However, the general public outside of the U.S. Southwest has very little experience with water insecurity, although the consequences of water insecurity affect other parts of the United States in significant ways. Amplified framing is more likely to influence uninformed respondents [31] or respondents with reduced exposure to or interest in an issue [23]. Therefore, citizens′ attitudes and beliefs about water insecurity are likely influenced by the way reporters frame this issue.

According to Druckman [46], the way elites frame an issue is a driving force for shaping public opinion. Coverage of elites in articles is of particular interest in this study as well, as this factor has an effect on the amplification intensity that news sources adopt over time. Elites are individuals in power relationships or positions of authority.

The source of unstructured text analyzed in this research offers researchers a body of content in which to explore the production of language. The issue of water insecurity is a case that can be studied to explore the patterns of interaction between online news sources and the effects of their frame choices. This research can generally apply to any large, unstructured dataset that is affected (or amplified) by framing. Therefore, it can be extended to other applications such as economics or political discourse.

Two other issues are offered here, to which the double subjectivity model may extend. Both are in the domain of government—U.S. health care reform and U.S. presidential elections with interdependencies with economic domain and intersections in disciplines in the social sciences and complex systems, as shown in Fig. 5.

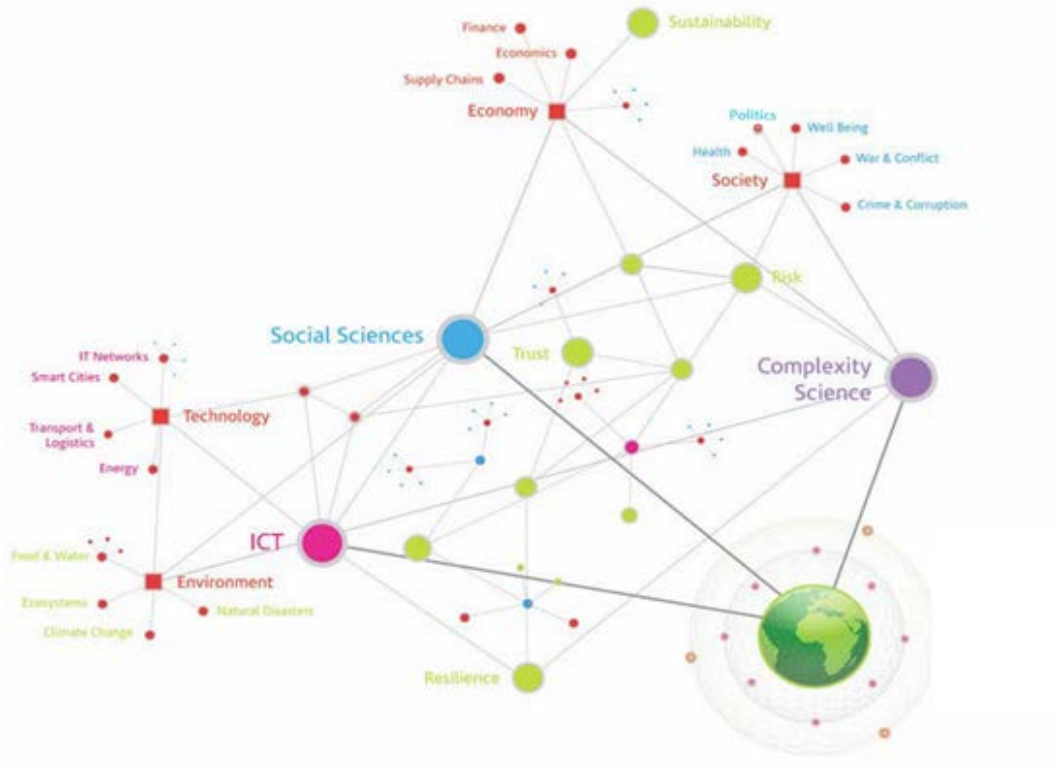


Fig. 5. Depiction of the interdependent domains and subtopics with intersections across disciplines. The image is important for showing the cumulative and cascade effects across interconnected domains. *Source: www.futurict.eu, Global Computing for our Complex World.*

F. The Case of U.S. Health Care Reform

The framing of online news media regarding the health care reform issue is an alternative contextual case for this research. Since passage of the Patient Protection and Affordable Care Act (ACA) in March 2010, the media coverage of this issue has been contentious and vast. According to Chong and Druckman [45]–[46], "politics is typically competitive, fought between parties or ideological factions, and issues that are debated are framed in opposing terms." As such—by its political nature and elites' cues to the

media—the issue of health care reform was instrumental in shifting public opinion [47]. Content analysis of 475 articles from seven top-circulating U.S. newspapers was conducted over a six-month period in 2010 [47], showing that the news media used two dominant frames—conflict and economic—for shaping public opinion about this critical issue.

The benefits that this research can lend to the U.S. health care reform issue include 1) an alternative to content analysis for handling the vast amounts of online news, 2) a tool for the automatic discovery of dominant frames, and 3) language and a visualization tool for analyzing dynamic changes in real time.

G. The Case of the U.S. Presidential Election

According to a report by Harvard Kennedy School′s Shorenstein Center on Media, Politics and Public Policy, the 2016 U.S. presidential election marks one of the most amplified and slanted elections ever [61]. The center reports that the preference of the election′s media amplification was decidedly negative. As scholars Robinson and Sheehan [62] noted over three decades ago, the news media seem to have taken some motherly advice and turned it upside down: ″If you don′t have anything bad to say about someone, don′t say anything at all.″ Strong negative amplified news has resulted in unintended consequences—namely, a growing distrust of government, elected officials, and basic facts [63]. One *New York Times* columnist went as far as saying that ″the

Internet is distorting our collective grasp on the truth″ [64]. In situations in which the flow of information is too vast to combine or synthesize, one has to resort to personal judgment 1) about the weight or critique that should be attached to the candidates; 2) the negative and positive elements surrounding the issue; and 3) the collective impact of the general conclusions on public perceptions, attitudes, and beliefs. Furthermore, content analysis is inappropriate because of the exhaustive inventory necessary to account for all online news articles, the need to filter out what is facts versus frame amplification, and the complexity inherent in the presidential election. Yet, over the last eight presidential elections, communication and social science experts have used content analysis to guide their understanding of the media framing effects on presidential elections, instead of developing tools that allow them to explore this issue in real time.

This research may lend the following benefits to analyses of the U.S. presidential election: 1) a tool for measuring amplification intensity over time and in real time, 2) a framework for exploring the social influence attributed to the double subjectivity with its primary property being the unique cumulative effect afforded in an online news network, 3) a model for exploring the dynamic forces at work when shifts occur in attitudes and beliefs, and 4) a visualization of the dynamic forces.

The term “framing” is used to describe cases in which online news sources use strategic devices for presenting prominent aspects and perspectives about an issue using certain keywords as well as stereotyped images and sentences for the purpose of conveying latent meanings about an issue [30]. Health care reform, water insecurity in the Southwest region of the United States, and U.S. presidential elections exemplify issues that are appropriate for study through the lens of double subjectivity that exists in

framing by online news sources because this issue is context specific, complex, and characterized by uncertainty. Newspapers regularly produce in-depth articles about these issues that are published online. However, the general public has very little experience with the underlying facts and the influence that media sources have for shaping attitudes and beliefs. Overly amplified framing is more likely to influence uninformed respondents [31] or respondents with reduced exposure to or interest in an issue [23]. Therefore, citizens′ attitudes and beliefs about these critical issues are likely influenced by the way reporters frame these issues. These sources of unstructured text offer researchers bodies of content in which to explore the production of language in the context of a socioenvironmental issue. The issues are case studies needed to explore the patterns of interaction between online news sources and the effects of their frame choices. This research can generally apply to any large, unstructured dataset that is affected (or amplified) by framing. It can, therefore, be extended to other applications such as economics.

H. Chapter Summary

Chapter 2 contains a review of related topics that provide insight into and understanding of the precursors and contributors to the computational treatment of unstructured text, communications basis, social influence, and culturally dynamic bodies of knowledge. A historical perspective contributes to an exploration of the evolution of the computational world for understanding pathways to the cognitive world. The review includes an examination of news frames as an instructional tool designed to shape readers' attitudes, beliefs, and values.

Chapter 3 of this research inquiry offers several improvements over traditional methods of discovering latent meanings and hidden structure in narrative text: First, unlike the traditional approach to processing unstructured texts by focusing on the characterization of form (syntactic) and its meaning (semantic meaning) [50], this method views news texts as organized, symbolic devices (frames) that act as signed vehicles (or cues) influencing perceptions about an issue that will interact with individuals′ existing beliefs (i.e., other senders or receivers of news articles). Second, by placing framing within computer science, this allows for rigorous processing of text representations that has the potential to extend bodies of research beyond the distribution of words and annotation. Third, this approach enhances current methods of text analysis and mining of narrative text by using machine learning to explore larger datasets and a plethora of issues, which go beyond what human experts and coders can currently evaluate using content analysis. Moreover, this new approach will allow for the automatic emergence of the frame from the text at the end of the analysis [65]–[66], using a combination of inductive and deductive iterative techniques.

Chapter 4 defines an online news issues network for exploring the network structure that comprises news sources and authors interacting under dynamic forces. Expressing double subjectivity in a network has potential to advance what is known about the production of digital language.

In Chapter 5, the double subjectivity social influence model is described for expressing the multidimensional aspects of news framing—and the amplification that it conveys—for transformational and maximal influence. The definition of the model, as

expressed in this chapter, involves the computational formulation of amplification intensity as a function in operation in the narrative text.

Chapter 6 contains the visual language for reading the network and exploring its dynamics using the model defined in Chapter 5.

Chapters 7, 8, and 9 comprise a summary of claims and recommendations for future research and practical applications.

I. Previous Publications

Publications that have directly contributed to this dissertation include:

- (In Review) Loretta H. Cheeks & Ashraf Gaffar (2017). Programmable Insight: A Visual Language for Expressing an Issues Network using Online News Frames. IEEE Transactions on Knowledge and Data Engineering (TKDE) 2017.
- Loretta H. Cheeks & Ashraf Gaffar (2016). A Social Influence Model for Exploring Double Subjectivity Through News Frames in Online News Sources. IEEE Intelligent Systems Conference (IntelliSys) 2017.
- Loretta H. Cheeks & Ashraf Gaffar (2016). Exploring Double Subjectivity Through News Frames in Online News Sources: A Network Approach. 2017 IEEE 11th International Conference on Semantic Computing (ICSC 2017).
- Loretta H. Cheeks, Tracy L. Stepien, Dara M. Wald & Ashraf Gaffar (2016). Discovering News Frames: An Approach for Exploring Text, Content, and Concepts in Online News Sources. *International Journal of Multimedia Data Engineering and Management* (*IJMDEM*).

- Loretta H. Cheeks, Tracy L. Stepien & Dara M. Wald (2016). Discovering News Frames: Exploring Text, Content, and Concepts in Online News Sources to Address Water Insecurity in the Southwest Region. 2016 IEEE 17th International Conference on Information Reuse and Integration.
- (In Preparation) Dara M. Wald & Loretta H. Cheeks (2017). Elites Framing: A Study Examining the Issue of Water Insecurity in the Southwest Region. Communication Research (CR).

II. Review of Literature

Scholarly research in computer science regarding the latent meanings in the associations between and among terms and documents to reveal relationships is found in literature related to text summarization, information retrieval, and text data mining. The earliest paper on text summarization is that of Hans Peter Luhn [15], which describes work being done at IBM in the 1950s. In his work, Luhn proposed that the frequency of a particular word provides a useful measure of its significance. Luhn′s contribution is the identification of the concept “term frequency” (TF), which says that it is possible to identify the significant terms just based on the TF calculated within that document and that the average, or entropy, of a term or group of terms can be used to rank how the terms are related to each other.

In the same year of the Luhn article publication, Baxendale [67] published work done at IBM that provides early insight on a particular feature helpful in finding salient parts of documents: the sentence position, which assumes that important sentences are located at the beginning or end of paragraphs. The following year, Maron and Kuhns [68] published ″On Relevance Probabilistic Indexing and Information Retrieval.″ This paper was the first on ranked retrieval and ranking of documents by their computed values of probability of relevance, which is important because two-valued indexing of documents could be replaced by weighted indexing, where the weights were to be interpreted as probabilities.

Edmundson [69] was the first to describe a system that produces document extracts. His primary contribution was the development of a typical structure for an

extractive summarization experiment that integrates features of word frequency and positional importance, influenced by the works of Luhn [15] and Baxendale [67].

In 1972, Karen Spärck Jones published in the *Journal of Documentation* a paper called ″A Statistical Interpretation of Term Specificity and Its Application in Retrieval″ [70]. Her contribution proved to be a giant leap in text summarization and information retrieval. She is credited with developing the measure of term specificity that later became known as inverse "document frequency" (IDF); it is based on counting the number of documents in the collection being searched that contain (or are indexed by) the term in question. The intuition was that a query term that occurs in many documents is not a good discriminator and should be given less weight than one that occurs in a few documents; the measure was a heuristic implementation of this intuition.

Critical to text summarization is the contribution of Salton et al. [16], credited with developing the vector space model (VSM). Extending the work of Luhn, Salton′s contribution for describing similarities (and dissimilarities) that are computed using both extracted words and cited data remains a principal concept upon which much of scholarly research leverages. According to Salton, a retrieval model represents documents, description features (such as index terms), queries, and the relationships within and across those sets.

Of Susan Dumais′ [25] research contributions, the one that aligns with this research study is her improvement to the VSM, known as latent semantic indexing (LSI). The development of LSI marks the first scholarly research for transforming a high-

dimensional VSM to associate terms within documents into a low-rank approximation to A_{mxn} to filter out noise and improve the detection of relevant documents. LSI uses singular value decomposition (SVD) to derive particular latent semantic structure models.

Other outstanding contributions to text data mining in the 1990s and beyond, with the advent of machine learning, include text representation [16] and models construction [66], [17], [19]; data dimensions reduction research in feature extraction [21], [71]; research on mining algorithm of text classification [19], [74], and clustering [72], [73]; and deep semantic mining based on natural language process [75], [76].

Frame discovery is motivated by framing theory, as it is known in communications research [24], which focuses on understanding the latent meanings of observable messages in their contexts and can provide important insight into how the presentation or framing of an issue affects the choices people make. Other disciplines have focused on framing. In linguistics research, similar approaches are described as "latent semantic analysis" (LSA; [39]) or social network analysis (SNA), which focuses on the importance of relationships among interacting units [26], [35].

III. Discovering Frames in Online News

A. Opinions and Social Influence

Whereas text data mining techniques and methods have given insight into first-level subjectivity, dynamic systems modeling provides insight into the double subjectivity that exists in narrative text. Models of social influence, cultural dynamics, and information diffusion (i.e., topic modeling, sentiment analysis, and opinion mining) are active areas of research [7], [1], [27]–[28].

The opinion models, such as the voter model, give insight into the spread and distribution of opinions [1]. It has been posited that social impact theory explains the impact of a social group on an agent as being dependent on the prominence of the social sources, their proximity, and source group mass (i.e., the number in the group) [6]. Political opinions and Axelrod′s cultural dynamics model behave similarly to these models that capture the interplay between selection and influence [2]–[5].

B. Networks

The ability to visualize vast amounts of data over time clearly creates a capacity for gaining insights about critical issues affecting society. At the most basic level, a network is a collection of points joined together in pairs [43]. The data points are called “nodes” (or “vertices”), and the pairs connections are called “edges” (or “links”). Here, in this manuscript, the terms nodes and edges are used. A network is a powerful way to represent patterns of connections or interactions between the parts of a system [43]. The issues network is best expressed using a network to understand the underlying global and

local structure of the network, the connections and interactions between nodes, the clusters of highly connected nodes, influence patterns, and migration paths.

IV. Discovering Frames in Online News

A. Introduction

The causal effect of media communication has been studied extensively, specifically the influence of the words and frames the media uses to influence public perceptions of social and environmental issues. However, it traditionally has been examined using content analysis, which was developed specifically to aid in the interpretation of social discourse or text for communication research [58], [49], [78]. Content analysis involves methodical evaluation and the systematic assignment of communication content to categories according to rules and the analysis of relationships involving those categories using statistical methods [79]–[80]. Current approaches to content analysis, however, require scholars and researchers to thoroughly examine documents in search of patterns in the text. This approach to text analysis is dependent on humans and limits the applicability of the analysis to large-scale unstructured text. Thus, a more effective tool for unlocking latent meanings found in unstructured text could enhance an understanding of online behaviors, responses to online advertising, and media influence on public perceptions.

Text mining, also known as "text data mining" [81]–[82], is a multidisciplinary field involving information retrieval, text analysis, information extraction, clustering, categorization, visualization, database technology, and machine learning. Text mining, coupled with an interest in understanding social influence and information diffusion for document summarization (i.e., topic modeling, sentiment analysis, and opinion mining), is an active area of research [27]–[28], [83]. Text mining, when combined with machine

learning algorithms, techniques, and methodologies, offers an added value to data integration tasks. Text mining highlights the similarities between heterogeneous sources and text features, which reduce uncertainty and risk exposure when performing the integration tasks.

The widespread availability of large data repositories, like those found in online news articles, creates an opportunity to develop new methods of text mining. These methods use machine learning algorithms and mathematical techniques to select, organize, and evaluate large quantitates of unstructured text. Although research on the use of frames and public attitudes in traditional news venues has been widely explored, the identification and analysis of frames in unstructured online news has received minimal attention.

Therefore, the purpose of this research is to test a new approach for discovering distinct frame patterns using a process that advances semisupervised machine learning for document clustering and classification. This research aligns favorably with concurrent efforts in machine learning and data mining that seek to discover novel patterns and latent relationships in unstructured text for deep learning [84]. This method has several improvements over traditional methods of text mining. First, unlike the traditional approach to processing unstructured texts by focusing on the characterization of form (syntactic) and its meaning (semantic meaning) [85], this method views news texts as organized symbolic devices (frames) that act as carriers or signed vehicles (or cues) influencing perceptions about an issue that will interact with individuals′ existing beliefs (i.e., other senders or receivers of news articles). Second, by placing framing within computer science, this approach allows rigorous processing of text representations that

has the potential to extend bodies of research beyond the bag-of-words model such as exposing pathways for bridging the bag-of-words model, clustering techniques, concept mapping, and linked data. Third, this technique enhances current methods of text analysis and mining by using machine learning to explore larger datasets and many more topics than human coders can currently evaluate. Moreover, this new approach will allow for the automatic emergence of the frame from the text at the end of the analysis [86], [66], using a combination of inductive and deductive iterative techniques.

This research inquiry has the aims of giving insight by 1) formulating a formal computer science definition for framing, 2) defining an approach for the discovery of five distinct patterns (referred to as signatures) that characterize prominent frames, and 3) proposing a process that advances machine learning for document classification that takes advantage of content analysis on a small number of documents for scaling up to meet the demands of large datasets.

B. Methods

In this section, data are described for the problem formulation, and the methods are described for discovering frames contained in unstructured data of online news article.

1) Dataset Description: The news data for this research were collected from Google News,[1] which is a news feed aggregator. After the noticeable water droughts started in the U.S. Southwest region, online news reporting on this issue became increasingly important given the uncertainties and potential risks to the affected states. The feature selection comprises four characteristics of a given article—namely 1) the news articles being published by online news sources, 2) the news source that generates

[1] http://news.google.com/

the articles, 3) the frequency of news publications by sources over time, and 4) the key features contained in the article.

The time period for the publication of the news articles was restricted to January 2006 to March 2015. A web data mining software program was developed using the Python programming language, which utilized a universal web browser bot for traversing the online frontier for gathering Uniform Resource Locator (URL) seeds for collecting online news articles that contained pertinent features needed for this research study. The URL seeds were used as input to the crawler.

A depth-first search algorithm was applied for each URL seed node for this study. Articles were gathered by limiting the search for keywords associated with water or subtopics of water consequences within Arizona, Colorado, California, and Nevada. News articles were collected (55 000 in total), and the articles were, then, stored in a database for undergoing data preprocessing that included removing errors and inconsistencies in order to improve the quality of data. For example, date patterns were standardized, duplicates were removed, and articles with missing critical values were removed (except in cases in which the author's name was not identified). After data preprocessing, 30 000 articles were deemed relevant for the news dataset, and 25 000 were deemed irrelevant. The web data mining process is graphically presented in Fig. 6. The web data mining program developed applied string, regular expressions, and tree matching techniques to find patterns and perform alignments. The data records were segmented into an association list, called "documents," in which an alignment of data items, called "properties," contained in the data records were produced for storing in a table in the database.

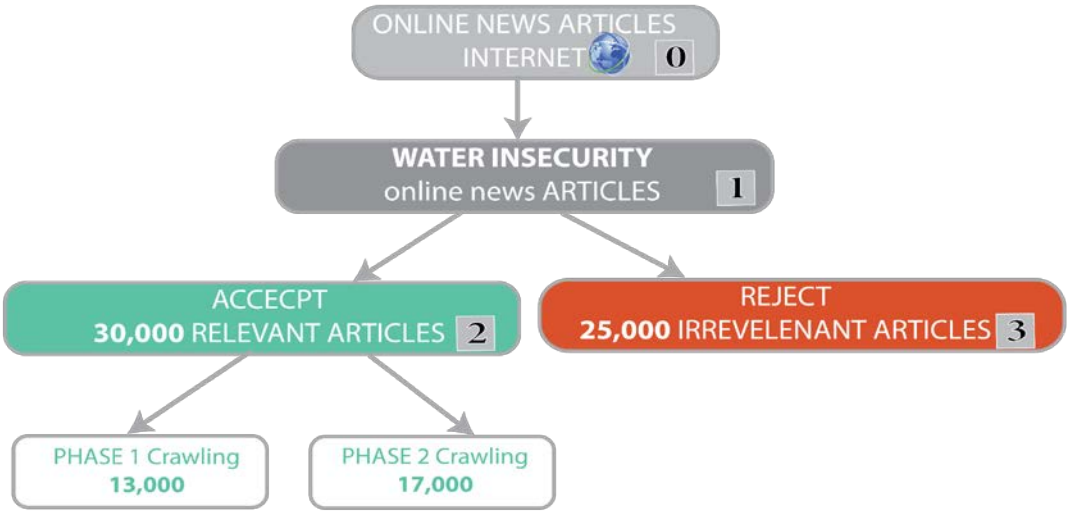


Fig. 6. Depiction of the process for the web data mining
Figure 6 16

2) Frame Discovery Methods: News frames constitute a central organizing idea that supplies a context and highlights salient features by selecting what to include, emphasizing or deemphasizing certain aspects of a news story, and elaborating or excluding particular perspectives, thereby shaping readers' interpretation of the issues [87]–[88]. Using this definition, I experimented with existing techniques and algorithms for machine learning.

Scholarship exists for text mining and machine learning for novel knowledge discovery of patterns as mentioned previously. However, this is the first known research to apply text mining and machine learning for discovering communication frames found in online news articles. Moreover, a discovery of frames may expose clues of influence that shed light on the intent of creating discourse about the issue central to the online news article and how adversarial news frames limit public understanding of

environmental issues, such as climate change, water insecurity, agricultural insecurity, health disparities, civil unrest, and death [89].

3) Content Analysis: To explore the relationship between online news source framing, a content analysis was performed on articles pertaining to water insecurity in the Southwestern region of the United States. Content analysis is a research method that involves extracting information from text by identifying characteristics or categories of content within a text [79]. Fig. 7 shows the process applied for the content analysis. Quantitative content analysis using statistical methods is used as a tool for evaluating numerical patterns (e.g., most of the items described this concept) and identifying relationships among the categories [79]. Content analysis is a research approach that uses a systematic application of rules and procedures to transform the content of communication (e.g., text, video, audio) into quantitative data that can be analyzed and interpreted. The term “coding” is the processing of classifying the unstructured text, and those performing the coding are called “coders.” Here, categories were identified in consultation with a subject matter expert; thereafter, *a priori* guidelines were identified to measure differences in content.

The data collection process occurred in phases. First, the crawler identified 13 000 articles that met the initial list of key search terms: water insecurity, water crisis and shortages, natural resource management water, natural disasters water, clean water, water rights, water policy, water river, water crisis, water doubt, water shortages, water desalination, water reclaim, and climate change.

Fig. 7. Content analysis process.

Quantitative content analysis [79] was used for calculating the sample size, which considers 1) the standard error and confidence interval of a given sample mean and 2) the estimated variance of the variable in the population. Assuming a 95% confidence level and a 90% minimal level of agreement, I identified a random sample of 372 articles [47]. Several of these articles were disallowed and counted as not relevant for this study, as they were found to be loosely related to the issue of water insecurity. The final sample included 316 articles, which was still sufficient for a 94.5% confidence level.

The content analysis involved the following steps. The overarching research problem was defined: “How is the issue of water insecurity in the Southwest region of the United States being framed in online news media over time?” Next, a list of variables of interest was developed related to the discourse around and framing of water insecurity in the Southwestern United States: 1) frame types (i.e., human interest, managerial, science, economic, conflict), 2) key actors quoted or cited, and 3) discourse and key actors over

time. The data collection protocol was designed based on these key variables. The protocol included a list of questions that SES researchers needed to address for each item of unstructured text (e.g., online news article). For example, one question included in the protocol was the following: “Does the article mention a science study?” Possible responses included “Yes” and “No.” There were a total of 44 questions designed to identify frame types, key actors, and discourse over time. All the questions addressed here were categorical with “Yes” = 1 and “No” = 0 response options.

Content analysis typically relies on the judgments of multiple coders. To ensure coding reliability, defined as “intercoder agreement” or “the extent to which independent coders evaluate a characteristic of a text and reach the same conclusion” [29], 90 articles were selected for coder training; These 90 articles were not included in the final sample. Three coders, including myself, participated in three training sessions of 30–60 minutes in length. Initial tests of intercoder agreement failed to produce reliable results; thus, the coding guide was modified, and coders participated in additional training. After making these changes, we were able to achieve > 85% agreement between the coder pairs. The content analysis process was laborious and took an extensive amount of time (9 months) to achieve coder reliability. The coding guide was revised multiple times, and questions were removed from the analysis if agreement could not be achieved regarding them. The responses to the content analysis survey questions were then put into SPSS (IBM SPSS Statistics for Windows, Version 23, Armonk, NY: IBM Corp.). Analysis showed five dominant frame types—human interest, managerial, science, economic, and conflict. These dominant frame types were established as the ground truth labels [57].

4) Problem Formalization of Frame: Let U_t be a universe of online news

documents at time t. Let $\{d_i\}, 1 \leq i \leq n$, denote a finite subset of documents from the universe of documents U_t, where n is the total number of documents being considered. Let $S_i = \{P_i, C_i\}$ be the set of document properties, where P_i is the source that produced the article and propagates a central organizing idea (i.e., framing) and C_i is the set of terms (i.e., features or content descriptors) from the news article with common words removed, which supplies a context and suggests what the issue is through the use of selection, emphasis, exclusion, cues, and elaboration [54]. For the issue of water insecurity, the standard communication frames were employed [56]—in particular, human interest, conflict, economic, managerial, and science. Let $f_1, f_2, f_3, f_4,$ and f_5 correspond to each of these frames, where each f_i is associated with a pair $\{x_i, y_i\}$ where $\boldsymbol{x_i}$ is the feature vector and $\boldsymbol{y_i}$ is the vector with the corresponding weights. Fig. 8 is a depiction of a problem formalization of frames. Machine Learning: In the search to discover framing within unstructured text, I considered machine learning to facilitate learning patterns. Machine learning algorithms and techniques have proven to be effective in selecting salient features for exposing emergent correlations and latent relationships. In the quest to understand how online news sources frame the issue of water insecurity, I established a machine learning process for discovering news frames. The overview of the process flow for discovering news frames is shown in Fig. 9.

The motive is to discover not only how meanings of subsequent terms, sentences, and paragraphs are related but also how the facts regarding these sentences are related. The news frame discovery process is composed of three main steps: data transformation, frame mapping, and classification.

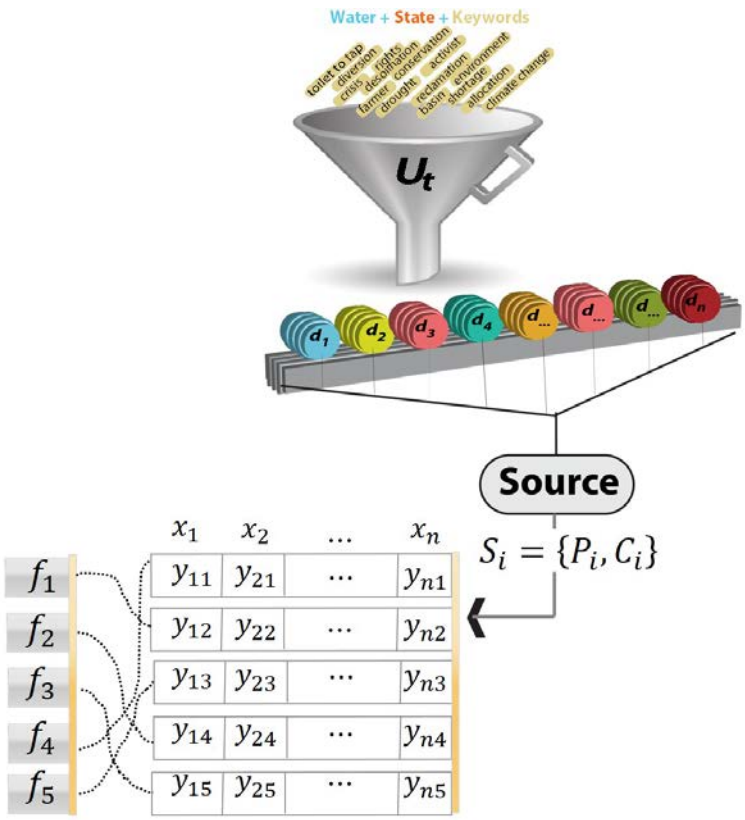


Fig. 8. Depiction of the problem formulation for the discovery of online news frames.

The performance of the algorithm depends on the quality of the vector space model for documents $\{d_i\}$. This research applied three methods for showing the quality of the vector space and evidencing machine-discovery accuracy: a) content analysis, b) machine learning techniques (Term Frequency-Inverse Document Frequency (TF-IDF), Non-negative Matrix Factorization (NMF), L2 Norm for comparing between vectors, and classification, which will be described in more detail in the Machine Learning section), and c) linked data for feature extraction of elite cues. This paper provides treatment for a)

and b) only to provide an approach for evaluating how salient features contained in online

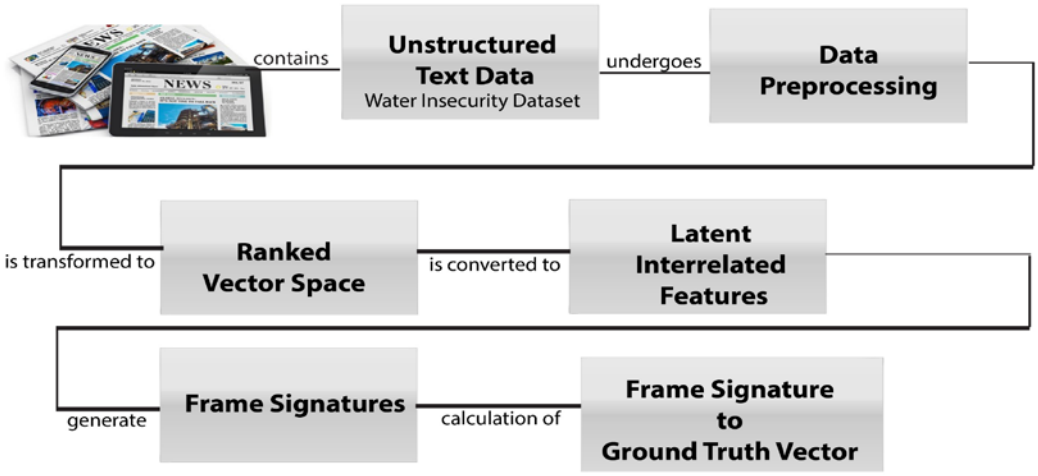


Fig. 9. Depiction of the overview of the process flow for discovering news frames using machine learning techniques and methods.

news articles are used for producing the frame about the issue covered in the article.

To explore the relationship between online news source framing, a content analysis was performed on articles pertaining to the water insecurity in the Southwestern Region of the United States. Content analysis is a research technique that involves extracting information from text by identifying characteristics or categories of content within a text [79]. Fig. 7 shows the process applied for the content analysis. Quantitative content analysis using statistical methods is used as a tool for evaluating numerical patterns (e.g., most of the items described this concept) and identifying relationships among the categories [79]. Content analysis is a research technique that uses a systematic application of rules and procedures to transform the content of communication (e.g., text, video, audio) into quantitative data that can be analyzed and interpreted. The term coding is the processing of classifying the unstructured text and those performing the coding are called coders. Here, categories in consultation with a subject matter expert were

identified, thereafter, *a priori* guidelines to measure differences in content were identified.

Because this research involves machine learning and the expectation to learn and, thereby, discover frames from a small dataset, the dataset was arbitrarily split into two subsets: about 66% of the data was used for a training set and 34% for a test set. Since the approach to text mining is novel, the accuracy of the classification methods underwent validation and testing by comparing the results of this machine learning approach with a more traditional approach to frame identification: content analysis.

5) Data Transformation: Learning from the article text requires the examination of the local coherence of the text through models that allow for feature analysis. The features can capture information related to the context of the article. The semantic meaning of a term may change depending on the context of usage. News features are helpful for capturing such information. In this study, two feature extraction models were applied for learning. The baseline model is TF-IDF, commonly known as bag-of-words, which is a weighting scheme used effectively to rank features based on association or co-occurrence and to build a vocabulary that will be used for deep learning [70], [16]. The TF-IDF baseline model is used as input to NMF [4], [30]. NMF has properties that are attractive for this class of problem—namely, dimensionality reduction—under the assumption of linearity and nonnegativity, clustering, and local minimum convergence [90]. A key goal of an NMF constrained model (e.g., a model

with only positive points) is to accurately identify latent relationships and to identify patterns that together explain the data as a linear combination of expression signatures.

The TF-IDF term "weighting scheme," for free text documents, applies to the evaluation of the specificity and, arguably, the importance of features contained in the article. TF-IDF transformation results in a VSM that enables document comparisons, term similarities, and term ranking as determined by the weight, which is referred to as either the "score" or "ranking." The primary assumption of the VSM is that documents that are "close together" in space are similar in meaning; therefore, in the VSM, each document is represented as a point in space (a vector in a vector space). The TF-IDF score (or ranking) of a term in a set of documents is calculated as follows:

$$w_i = tf_{dc} \times idf_c, \text{ and} \qquad (3)$$

$$tf_{dc} = \frac{n_{dc}}{m_d},$$

$$idf_c = \log \frac{\{d_i\}}{1 + |\{d_i \in \{d_i\} : c_i \in d_i\}|}$$

where n_{dc} is the frequency of term c_i in document d_i, m_d is the maximum frequency of any term in document d_i, $\{d_i\}$ is the total number of documents in the corpus, and $|\{d \in \{d_i\} : c_i \in d_i\}|$ is the number of documents in which t_i appears. Without the term in the corpus, division by zero will occur; therefore, the denominator is adjusted by adding 1. The weights w_i show the importance of the terms in each document. Documents were represented using a bigram of features. The interest is to retain only those features that occur in at least 5 documents in the whole training set. The VSM produced through the calculation of TF-IDF is a high-dimensional matrix, $\{d\} \in \mathbb{R}^{d \times c}$.

For the news article dataset, the TF-IDF data matrix w_i was used as input to the NMF matrix A that represents $d_1, \cdots d_n$ documents, such that each column corresponds to the features vector $\boldsymbol{x_i}$ from frame f_i. The sparse NMF extension by Hoyer [91] is applied, using SVD initialization (i.e., nonnegative double singular value decomposition [NNDSVD]) for addressing the convergence problem, which is common in many clustering algorithms (e.g., K-means or latent Dirichlet allocation [LDA]) [6]. The nonnegativity constraints in NMF result in the unsupervised selection of sparse bases that can be linearly combined to reconstruct the original data for learning parts of the news articles [92]. NMF of a matrix A decomposes it into two matrices W and H such that

$$A = WH. \tag{4}$$

If the dimension of A is $d \times c$, then W has dimensions $d \times r$, and H has dimensions $r \times c$, where r is the dimensionality (i.e., rank). Here, r is defined to be 5, which corresponds to the five dominant frames (i.e., human interest, managerial, science, economic, conflict) used by online news sources. In NMF, the goal is to factor A into r components for representing these frames. The H matrix denotes the five distinct signatures (patterns) where the columns are the important features for the five distinct frames.

This factorization is a constrained nonconvex optimization problem with the cost function equal to

$$\| A - WH^T \|_2^2, \tag{5}$$

such that $W = [w_{d\times r}] \geq 0,\ H = [h_{r\times c}] \geq 0.$

The cost function is convex only in W or H, but not convex in both together [10]. The gradient descent contribution of Hoyer [91] extended Seung and Lee's [30] iterative update rule for overcoming the nonuniqueness starting point that results without a global solution for the algorithm while presenting an improved convergence to a local minima of the cost function.

When training this model, TF-IDF vectorization preprocessing is performed that includes 1) selecting an n-gram range of two for denoting an interest in bigram features, 2) ignoring terms that have a document frequency strictly lower than four, 3) removing from the vocabulary common words, called *stopwords* in text mining, that would appear to be less significant in selecting document features, 4) tokenization for breaking up given character sequences found in the document into meaningful words, and 5) normalizing for different document lengths using the L_2 norm. NMF uses 1) a NNDSVD seed, a method designed to enhance the initialization stage of NMF [6], 2) the rank of the decomposition equaling 5, and 3) an updated iteration of 50 times before timing out.

6) Frame Mapping: The frame mapping function is a four-step process that comprises 1) sorting, 2) aligning, 3) measuring, and 4) mapping. The NMF part-based representation matrix H, provides an efficient manner to discover hidden structure and latent relationships within data [11]. Each frame signature is mapped to a document label that is applied during the content analysis categorization of dominant frames (i.e., human interest, conflict, economic, managerial, and science). The frame discovery process that utilizes the content analysis is the ground truth for the low-rank approximation that results from the documents reconstructed using the NMF H matrix. The feature names

were preserved during the transformation.

The frame mapping process takes as input the H matrix, the best rank-r approximation of the original matrix A. To create a single frame signature, one should start the transformation using the TF-IDF matrix D that contains the content analysis label for the dominant frame under observation, and then calculate the NMF H. Here, I call this H matrix the "single signature," s_1, which means that its rank-$r = 1$. The single signature H matrix should be differentiated from the entire training dataset to derive the H matrix where the rank-$r = 5$. The sort is performed using the s_1 H matrix according to decreasing ranked values against each frame signature $f_1, f_2, f_3, f_4,$ and f_5, as shown in Fig. 10.

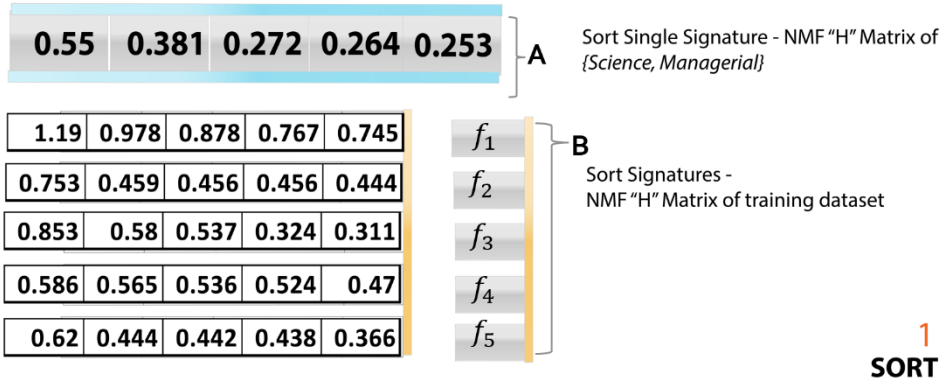


Fig. 10. Both H matrices sorted.

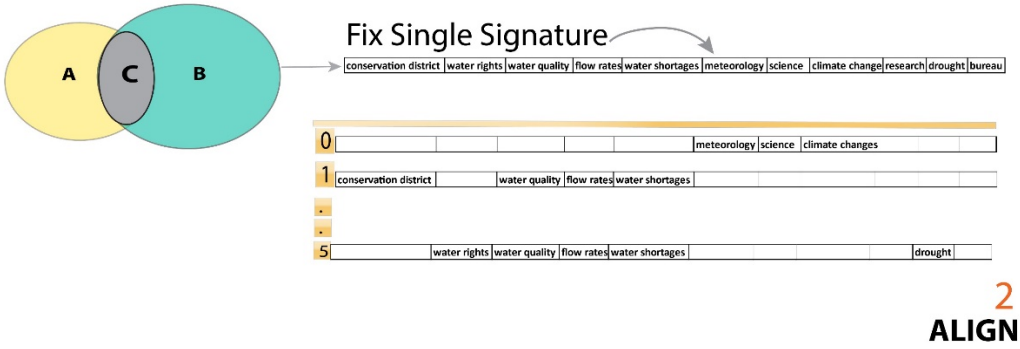


Fig. 11. Align the single signature to the relevant features in preparation for distance measurement.

The alignment is achieved by 1) iterating each row in $f_1, f_2, f_3, f_4,$ and f_5, 2) calculating the intersection between the single signature and $f_1, f_2, f_3, f_4,$ and f_5, 3) building a feature vocabulary from the resulting intersection for establishing a master single signature, and 4) using $f_1, f_2, f_3, f_4,$ and f_5 to build a subset of the H matrix for extracting features relevant to that row. This must be done to perform a one-to-one measurement of features, see Fig. 12.

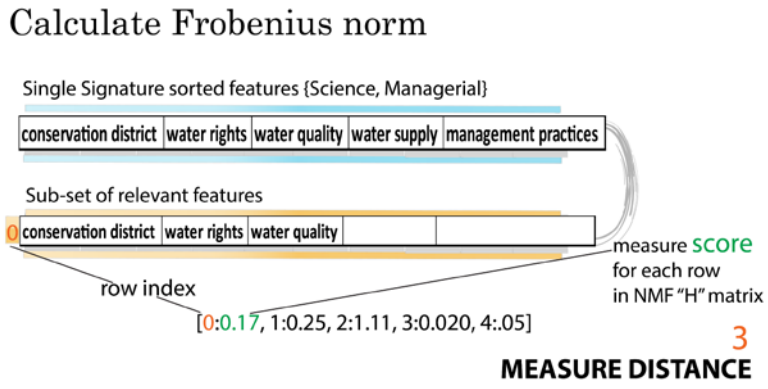


Fig. 12. Calculate the distance measure.

The next step is to calculate the similarities of the s_1-sorted H matrix compared with the subset feature matrix that contains the extracted relevant aligned matrix using the L_2 norm for discovering the frame for each document. This yields a list that contains the row index and value that correspond to $f_1, f_2, f_3, f_4,$ or f_5. The minimization of $(s_1 - \{f_1, f_2, f_3, f_4, f_5\})$ gives the frame that is most similar to the single signature, as shown in Fig. 13

The next step is to use the row index to map the s_1 to $f_1, f_2, f_3, f_4,$ or f_5. Thereby, the document is recognized as the frame candidate and, as such, receives the class label of the frame signature selected (see Fig. 13).

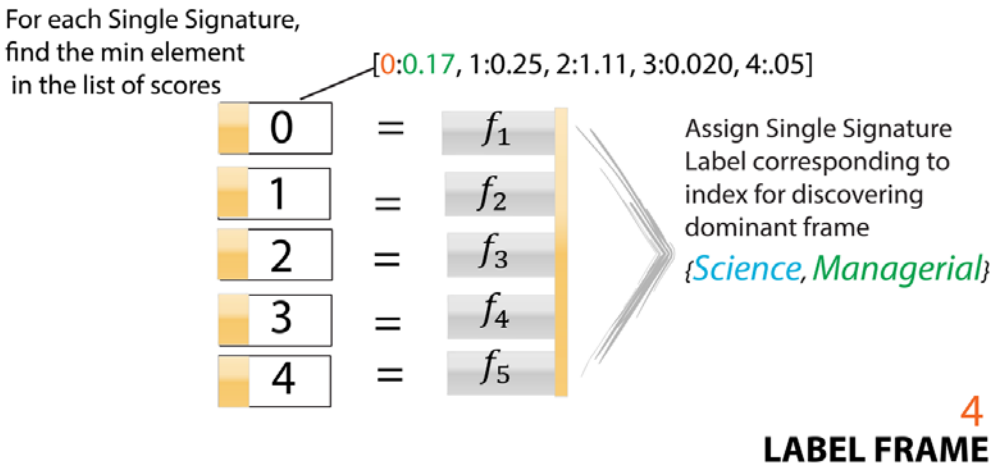


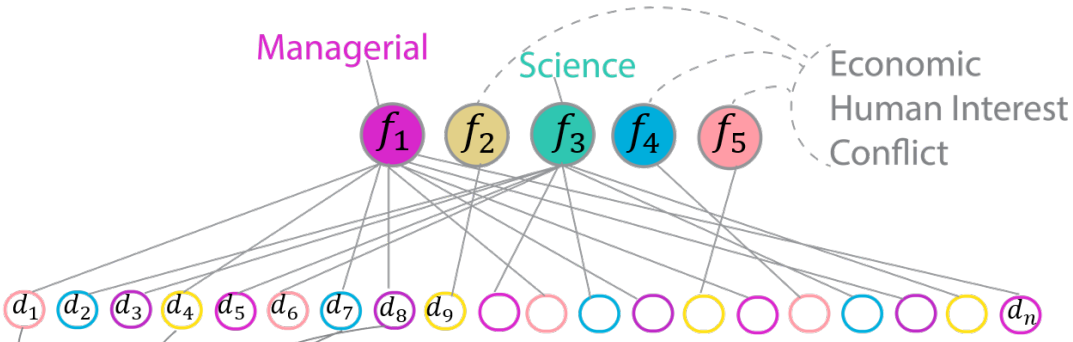
Fig.13. Depiction of the assignment of dominant class label to training documents.

C. Evaluation and Results

To examine the effectiveness of the news frame discovery approach trained by a small dataset size in the context of the issue of water insecurity, the next step is to perform the experiment with respect to varying parameters for NMF and the use of anchoring features found to be prominent in the learned vocabulary. The basic premise

for applying the ground truth is that it acts as a quality measure for comparing the content analysis labels for a subset of the data to algorithm-generated hidden structures to understand similarities and boundary differences [29]. The ground truth label is a method of validation for the machine learning prediction.

The frame mapping function successfully maps dominant frames to $f_1, f_2, f_3, f_4,$ or f_5. To prepare for the prediction, each document in the training dataset must be assigned a class label. The assignment can be determined by calculating the distance measure of the TF-IDF matrix that is a derivation of the training set against $f_1, f_2, f_3, f_4,$ or f_5, as shown in Fig. 14 To arrive at the same matrix dimensions for performing the distance measurements, a process similar to that described in the frame



	water supply	water resources	global warming	monitoring technic	water rights	water supplies	desalination plant	water crisis	atmospheric resear	water resources	states water	state water	surface water	climate changes	department water	water authority	water shortages	class label
d1	0.089	0.077	0.016	0.016	0.101	0.096	0.091	0.099	0.061	0.076	0.093	0.067	0.069	0.064	0.275	0.085	0.070	Managerial
d2	-0.082	-0.015	0.003	0.003	-0.007	-0.017	0.085	-0.096	0.084	-0.065	-0.080	-0.019	0.020	0.040	0.493	0.178	0.107	Science
d3	0.000	0.027	0.024	0.024	-0.028	0.072	-0.088	-0.005	-0.077	0.006	-0.024	0.001	0.070	0.070	-0.396	-0.141	-0.106	Science
d4	0.004	-0.011	0.008	0.008	-0.038	0.168	-0.016	0.001	-0.105	-0.012	-0.022	-0.006	0.042	-0.072	-0.133	-0.083	-0.010	Managerial
d5	-0.037	0.000	-0.005	-0.005	0.006	0.146	-0.043	-0.037	0.011	-0.015	-0.032	0.095	0.042	0.197	-0.157	-0.053	-0.046	Science
d6	-0.002	0.016	0.016	0.016	-0.015	-0.103	-0.002	-0.008	0.011	0.016	0.009	-0.027	0.021	-0.091	0.093	0.041	0.021	Science
d7	0.035	-0.019	0.015	0.015	0.004	0.263	0.031	-0.012	0.096	-0.016	0.019	-0.140	0.093	-0.055	-0.090	-0.023	-0.014	Managerial
d8	-0.024	-0.007	-0.154	-0.154	-0.030	0.696	0.025	-0.015	0.010	-0.009	-0.011	-0.035	0.021	-0.026	0.101	0.051	0.008	Managerial
d9	-0.021	-0.006	0.065	0.065	0.037	-0.318	0.005	-0.013	0.017	-0.012	-0.028	0.034	0.054	0.000	-0.022	-0.043	0.019	Unknown
dn	0.009	-0.017	0.312	0.312	-0.029	0.283	0.019	-0.002	0.023	-0.026	-0.047	-0.051	-0.011	-0.028	0.097	0.046	0.012	Unknown

Fig. 14. This diagram shows the outcome to the machine learning frame discovery mapping process using a matrix. The matrix comprises documents in relationship to the bigrams that receive a frame choice class label.

mapping function should be conducted. This step is a form of reverse engineering, the part-based representation of H matrix back to the d_i corpus.

Algorithm

input: **H** (frame signatures, $r \times c$), $\boldsymbol{w_i}$ (TF-IDF ranked vector space, $d \times c$)
result: a set of d_i −document id, frame class label and f_i −frame type
1: **for H** = $f_1,\ f_2,\ f_3,\ f_4,\ f_5$ **do**
2: sort each row in H
3: **for all** $\{d_i\}$ in $\boldsymbol{w_i}$ TF-IDF vector space **do**
4: **if** c_i is in $\boldsymbol{w_i}$ **then**
5: $L_i \leftarrow$ align c_i in new list
6: **end if**
7: **compute** L_2norm of $f_i - L_i$
8: choose *min* of $f_i - L_i$ results
9: **assign** discovered frame class label for $f_1,\ f_2,\ f_3,\ f_4,\ f_5$ to d_i
10: **end for**
11: **end for**

This process is designed to handle five distinct frame signatures; the experiment is trained on the science and managerial signatures. The frame discovery process is tested by 1) observing a news article test document d_i, and 2) deciding the most similar frame using the aforementioned frame mapping function. The process is trained on a subset of news articles (or documents) categorized as science frame during the content analysis, and the frame signature is tested by searching each document contained in the test dataset against the frame signature. The closest frame signature to the document is marked by its

machine learning frame type: science or managerial. The test data frame type is checked against the ground truth class label assigned to the document during the content analysis. Similarly, this training and testing is done for the managerial frame signature. In the case when the frame signature measurement is the same—a tie—make an arbitrary selection; thereafter, apply the selected frame signature class label.

The initial experiment contradicted the research study expectations. This finding led to the retraining of the model using a subset of learned vocabulary for anchoring the frame signatures (frame types) with the original training dataset and using it as a feedback loop for improved frame classification with the original test dataset. The holdout method was used for evaluating the process for frame discovery: one-third of the data reserved for testing and two-thirds of the data reserved for training. The performance of the experiment is reported by using standard measures of recall, precision, recall, and F1-score, as shown in Fig. 15 [7]. Thus, one interpretation of the results is that the constructed model prediction of science and managerial frames has potential for extending traditional content analysis, thereby allowing for the analysis of content on large datasets. These findings expose efficiency gains and an interesting correlation between the intercoder agreement and the machine learning of the system.

	PRECISION	RECALL	F1-SCORE
CLASS 0 (SCIENCE)	78%	82%	80%
CLASS 1 (MANAGERIAL)	75%	74%	75%

Fig. 15. News frame discovery results.

V. Online News Issues Network

A. Introduction

The source of unstructured text analyzed in this research offers researchers a body of content in which to explore the production of language. The issue of water insecurity is a case that can be studied to explore the patterns of interaction between online news sources and the effects of their frame choices. This research can generally apply to any large, unstructured dataset that is affected (or amplified) by framing. It can, therefore, be extended to other applications such as economics or political discourse.

The aims for constructing a news frame issues network using water insecurity as a contextual case are 1) to gain a deep understanding of how framing advances novel knowledge discovery with sensitivity to double subjectivity as a factor in the production of language in text narratives and 2 to explicitly examine the temporal dynamics of structural changes that may exist within the network. The overarching research problem is the following: ″How is the issue of water insecurity in the Southwest region of the United States being framed in online news media over time?″

B. Methods

Construction of Issues Networks for Online News: The first step in the construction of an issues network for news frames is to define the nodes. The issues network is a collection of news sources and article nodes of interest joined together in pairs by edge connections. In this research, the news articles being published by online news sources and the news sources that generate the articles served as the nodes. The boundary of the network in the study was determined to be water or subtopics of water

consequences within Arizona, Colorado, California, and Nevada and the online news sources who produce articles on this issue. These states are included in the map in Fig. 16 which shows states that are in drought. It is the reasoning here that news sources play a critical role in the propagation of frame amplification about issues, and amplification in online news drives a wedge between evidence and beliefs [36].

Since water is a fundamental resource affecting all aspects of life, one can expect that the number of articles being produced on the issue is an indicator of its prominence

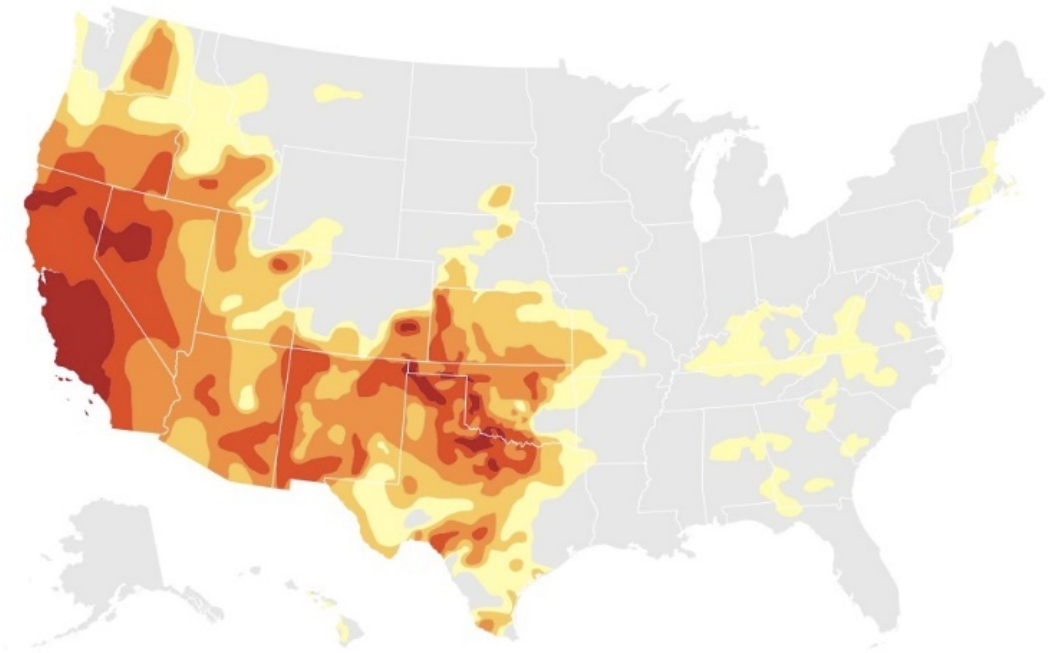


Fig.16. Drought severity impacting the United States. *Source: National Climatic Data Center, National Drought Mitigation Center, U.S. Department of Agriculture, National Oceanic and Atmospheric Administration.*

and the framing choices of the news sources. The framing of water issues is key for driving policy and decision-making, in addition to shaping attitudes and values. Such framing also indicates areas of amplification intensity. Connections are made between news sources through the capture of framing adoption over time as it happens during

migration. Migration pathways can be understood as carriers and adopters of online news frames.

Framing of critical socioenvironmental issues in online news is a complex and dynamic social network, which can be studied as a graph, as it is here. Graph analysis has become important in understanding the dynamic process in the production of news frames. Graphs allow for the visualization of the social network, showing interdependency of actors (or nodes) in terms of the social relationships such as friendship, kinship, or financial exchange [33]–[34]. Graphs represent objects where order and disorder coexist. Graphs serve well for showing social interactions, influence, migration paths, and framing effects. For instance, a graph may expose which news sources hold central positions that function as points of prominence, control, and stability and may show edges that act as highways for lead relationships and exchanges of news frames. These intertwined dynamics, coupled with the vast amounts of online news being produced daily, make graphs an important tool for visualization and analysis of information flows pertaining to critical issues.

The overarching research problem is expressed through the construction of an issues network, and the model of a news frame issues network is presented in this paper as a graph $G = ({S_g}^t, {A_h}^t, {E_r}^t)$ consisting of 1) S_g that denotes a set of source nodes containing $g = \{s_{1,}\, s_2, \cdots, s_g\}$ elements and 2) A_h as a set of article nodes containing $h = \{a_{1,}\, a_2, \cdots, a_h\}$ elements that operate within the time window t.

In general, t indicates one discrete time step of a reporting period. Let $S_g = \{C_i, F_i\}$ be the set of source properties, where 1) C_i is the set of terms (i.e., subissue keywords or content descriptors) and b2F_i is the dominant frame choice (identity) that the news source uses.

Each relation has a corresponding set of edge connections, $E_r{}^t$, directed/undirected as edge elements. The issues network subscript $r = 3$, corresponds to the total number of relations. The edge set represents the communication channels between node pairs. A_h are children of S_g; therefore, they may inherit the properties of the parent news source, as shown in Fig. 17.

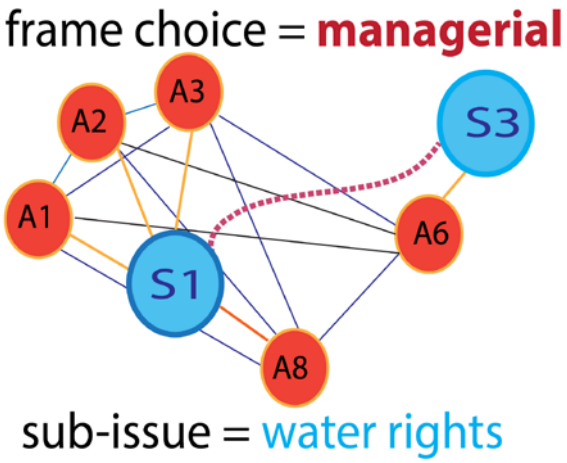


Fig. 17. A subset of the news frame issues network with two news source nodes with connecting articles using the same subissue and the same frame choice. The blue circles around notation S_g show the online news sources node that produced the article and propagates a central organizing idea (i.e., framing), orange circles around A_h show an online news articles node, black lines show edge connections among neighboring articles on the similar issue, golden lines show edge connections of news sources who produce articles covering similar issue, and the red line shows a managerial news frame.

The relations (or rules) for E_r edge connections are as follows:

$\boldsymbol{R_1}$. A & A (articles to articles). A nondirectional edge connection is constructed for E_1 when $A_h(a) = \{a: (a_1, a_2, \cdots, a_h) \in E_1\}$ is the set of neighbors of $\boldsymbol{a_h}$ such that $\boldsymbol{a_h}$ represents articles with a similar subissue.

$\boldsymbol{R_2}$. A & S (articles produced by the same source). A directional edge connection is constructed for $E_2$when $S_g \in E_2 \rightarrow A_h(a) = \{a: (a_1, a_2, \cdots, a_h) \in E_2\}$ is the source (or producer) of the set of neighbors of $\boldsymbol{a_h}$ articles, where $\boldsymbol{a_h}$ represents articles with similar subissues.

$\boldsymbol{R_3}$. S & F (sources with subissue to subissue with the same frame choice). A nondirectional edge connection is constructed for $E_3$when $S_g(s) = \{s: (s_1, s_2, \cdots, s_g) \in E_3\}$ is the set of neighbors of $\boldsymbol{s_g}$ such that $\boldsymbol{s_g}$ shares the same dominant frame choice.

Problem: Given $G = ({S_g}^t, {A_h}^t, {E_r}^t)$ about I issue, construct a graph for determining how frames are produced in online news media over time.

VI. Double Subjectivity Social Influence Model

A. Introduction

This research includes a model for exploring frames as instruments with online news articles, a form of narrative text, to expose a deep level of expressivity—what is referred to in this paper as double subjectivity. The double subjectivity model [50] is presented for gaining programmable insight into the hidden structures in text and understanding influence patterns. The double subjectivity method expands Axelrod's [7] and Klienberg's [28] research on cultural dynamics and social influence for constructing an issues network representation and designing a visual language for exploring spatiotemporal dynamics. The issues network is exemplar of a complex dynamic system whereby migration paths may be visualized. The migration path is defined as the movement of frame choices in articles across a network to establish a new bridge for information exchange. The way that the issues network grows (gain node connections) and shrinks (loss node connections) are discussed in a subsequent section in this manuscript.

The aim of the research is to demonstrate news framing as a step toward deep learning about the production of language that may inform scholars about pathways that offer new approaches that advance cognitive computation for improved intelligent behavior in machines.

B. Model Setup

To model the news frame issues network, I formulated a new model for information feedback based on a social system. The new framework is within the tradition of previous frameworks, models [1], [7], [28], [38] for studying the evolution of personal

position, social influence of multiple dimensions, and forces that cause cultural shifts. Axelrod's model [7] is built on two simple assumptions: 1) selection, a phenomenon in which people are more likely to interact with those who are more similar to them, and/or to be more receptive to influence from those who are similar and 2) influence, a term that describes the case in which interactions tend to cause similarity among interacting actors. The model in the present research promotes the additional emphasis of amplification that emerges as double subjectivity in narrative text. Although the behavior of the extended model is similar to that in Axelrod's [7], in decisions regarding how to interact in the network, this model differs in that insight is given into the endogenous and exogenous forces that may be at play in dynamic systems. Therefore, the double subjectivity in the social influence model integrates cultural dynamics extensions [28] with the CCM [38]. The CCM is useful for describing nonlinear relationships such as those found in narrative text. In the model introduced in this research, the particular focus is on the different forces at work in the amplified slant employed in news frames as it allows for the degrees of freedom whereby alternative pathways may be realized. The CCM allows for understanding the different forces at work when quantifying framing amplification and the migration path.

Conceptually, the switch from linear to nonlinear relationships involves taking into account not only a news source's capacity to select (or affect), but also another news source's capacity to be influenced (or affected). Thus, an important aspect of selection and influence is the characterization of dynamic systems, not just by their properties, but also by their capacities.

Consider this contextual example. When a news event occurs that impacts society, it, thereby, becomes an issue. It is partly defined by its properties, such as the absolute baseline facts, as well as being in a certain state, like the event having taken place on a certain day. This same news framing, however, has the capacity to intensify its amplification that may have the causal effect of a revolt or eruption by the people, an exacerbation of sentiment about a group of people, or the viral sensation of the news frame. This can happen through interactions with news sources that have the capacity to exercise strong amplification and to simultaneously weaken the agency of the people, group, or opposing view.

Properties are always reality based, since at any given point in time the facts—absolute baseline—are either true or false. Hence, facts are logical, formal, and rigid, but the causal capacity of news when exercising double subjectivity has a cumulative effect, which acts as a fabrication of what is real without being actual (or absolute). Unlike the limitations that facts in online news offer, amplification intensity within the construct of double subjectivity allows for the freedom to change dimensions through the parameterization of different forces—frame identity, justice, and amplification intensity—and distributions of sudden changes. The end effect is the perception of reality about an issue in online news that may offer new structure formations (or pathways) that lead to alternative views for shaping attitudes and beliefs. Stated differently, that which was only potential becomes a reality when changes in the different forces reach a certain critical threshold and a sudden jump happens. At this point, changes in beliefs can be tracked with the model, even if the news framing amplification mechanism morphs the facts into something far from the truth.

In the CCM, this possibility space is “the response surface space and control space,” which represent all possible states in the system. In this research, the relationship among frame identity, justice (or a sense of fairness), and amplification intensity are acknowledged through the cusp catastrophe differential calculus formula, which allows the measurement of the rapidity or slowness with which forces can change. In the geometric approach to the calculus, each degree of freedom becomes one dimension of a possibility space. The space of possible states in the system allows for the differential relations between them to determine a certain distribution of stable states around points or loops of attractions. They have a fractal dimension (intermediate between one and two) and are referred to as “chaotic attractors” [37]. Thom’s CCM [38] shows specific transitions of the forces, which have the tendency to cycle through the same set of states over and over.

The integration of cultural dynamics and the CCM for defining the new double subjectivity social influence model may shed insight into deep levels of expressivity in narrative text where amplification offers alternative pathways and possibilities. This remains an emerging research topic for understanding the production of language that may help in synthesizing vast amounts of unstructured text and learning online social behaviors.

In this dissertation, I explore the process that news sources use when producing online news articles. The main factors considered in this production are the facts, amplification intensity, and signals that derive from neighboring news sources. The signals that derive from news sources are likened to the reporter-relative or a floating

baseline as mentioned previously. In the process of time, the news sources' absolute baselines will undergo adjustments as the amplification intensity changes.

C. Double Subjectivity Social Influence Model

The double subjectivity social influence model integrates cultural dynamics extensions [7], [28] with the CCM [38], [40] for modeling double subjectivity in narrative text. The CCM is useful for describing nonlinear relationships such as those found in narrative text. The CCM allows for understanding the forces at work when quantifying amplification and the migration path.

Online news sources tend to produce articles about an issue I of interest to society, such as water insecurity, the state of the economy, or affordability of health care. Additionally, news sources may receive signals β from other news sources on the importance of the issue, based on observed increase in the number of articles other neighboring news sources produce. They receive facts f about an issue. Each news source, then, selects an amplification strategy α using as the basis of its calculation the amplification intensity. More specifically, each news source S_g with interest in I issue at time t generates news articles

$$I(\mathrm{g}) = \alpha_g + \beta_g(f) \tag{6}$$

where α_g and β_g are the coefficients of the news source while f represents the facts received. Furthermore, it is supposed that α_g corresponds to the endogenous propensity of the news source nodes to express its own amplification (or double subjectivity), whereas β_g represents the exogenous force in the network. In the first interaction, set $\alpha_g = 0$,

denotes amplification intensity; this value will be calculated upon interacting with neighboring nodes and should change to give shape to the news source's absolute baseline. Time evolves in discrete steps t = 0, 1, 2,⋯, and $S_g(t)$ denotes the mass count on S_g at time t. The maximization of $\{S_g\}$ gives the subissue most important, M_i—a count of the largest cluster (the mass) of nodes reporting on a water subissue. This initialization of α_g and β_g provides access to new nodes to enter the overall conversation about water consequences by linking to the most prominent mass, thereby, leveraging facts and the neighboring nodes of the relative baseline. This is similar to the operation observed in the hidden Markov model (HMM), whereby the news sources' absolute baseline is hidden with adjustments made over time. In contrast, the relative baseline is likened to the observations one can make in the HMM for getting a sense of the state [13].

Problem: Given $\{S_g\}$ interactions within graph G_t over time, determine the migration path and the nonlinear, causal linkages about $I(\mathrm{g})$.

At the start of the process, each news source $g \in S$ has a nonnegative node mass associated with it, corresponding to the fraction of the news sources that initially reports (i.e., through signaling) on similar subissues about water insecurity. The dynamic system allows for each news source to switch frame choice, $F(e)$, and subissue interest, thereby enabling random selection of news source interaction. Also, news sources are susceptible to being influenced when they interact with neighboring news sources that share similar amplification intensity. The full state space may be calculated by counting the news sources as expressed as the mass vector $S_g(t)$.

D. Defining Amplification Intensity for Expressing Double Subjectivity

The quantification of amplification intensity is situated in terms of the mathematical catastrophe theory [38]. Catastrophe theory is a branch of nonlinear dynamic systems theory that originated with the work of the mathematician Rene´ Thom [38] to help explain biological morphogenesis as one of the great mysteries confronting mathematical biology. A key property in catastrophe theory is that the system under study is driven toward an equilibrium through its use of gradient descent or potential function for seemingly automatic guidance (i.e., through the law of attraction) occurring in the system, which is important in research on social influence, particularly when considering the property of convergence and stable states. The cusp model is the most well-known and the simplest model of catastrophe theory, positing that nonlinear transition within a system from one state to another is guided by two controlling variables: asymmetry and the bifurcation factor.

1) Base Cusp Catastrophe Model: The operating principle of the catastrophe theory and CCM is clearly described by Thom [38] and an abbreviated description of the model operation is given here. The CCM can be formalized by potential or gradient structures; a potential function $V(X)$ is a function of both the system behavior X and the control parameter(s) M and N. In mathematical terminology, the CCM involves one state variable X and two control variables M and N. Let X be the behavior variable for the detection of frame amplification, and let the potential function F reflect the ability of the framing of the issue to maintain stable operation. The standard form of the potential function V can be expressed as follows [97]:

$$V(X) = x^4 + Mx^2 + Nz$$

The equilibrium curved surface is given by Eq (x):

$$\frac{\partial F}{\partial x} = 4X^3 + 2MX + N = 0$$

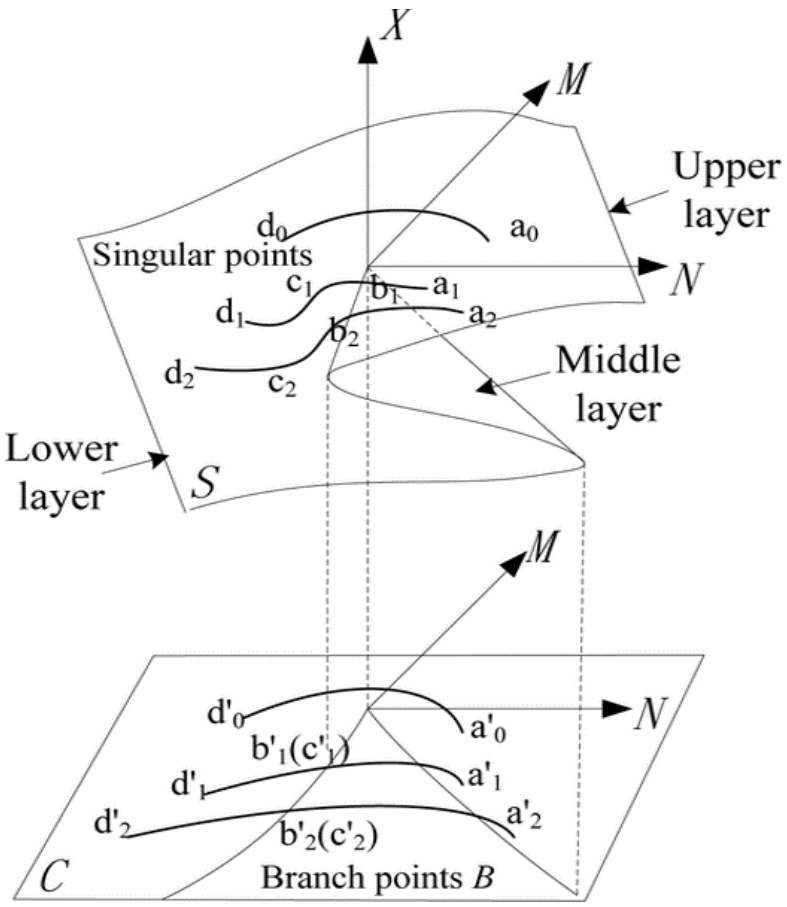


Fig. 18. Double Subjectivity Social Influence Model with consideration for the coefficient amplification intensity. The two planes shows S is the equilibrium surface (or possibility space) and C is the $M - N$ control surface.

The curved surface is called the “catastrophe manifold”; it represents the conditions met by the equilibrium state of the system, and its corresponding point is called a “critical point” in mathematics [96]. The entirety of the points on this surface (X, M, N) are referred to as the equilibrium points of the potential function V. The equation of singular points can be written as follows:

$$12X^2 + 2M = 0$$

If we project singular points onto the control surface $M - N$, the branch points B can be expressed as follows:

$$8M^3 + 27N^3 = 0$$

Changes in the control variables M and N essentially cause stable changes in the behavior variable X. A discontinuous catastrophe will occur only when the controlling points (M, N) jump over a branch point B. The points located on the two branches of B are catastrophe points (i.e., singularity points [97]), as shown in Fig 17.

2) Model Variables Defined: Frames are a characteristic of human cognition structures that operate simultaneously in human minds and texts. These latent structures of subjective interpretation are often produced and transmitted through the lens of power relations, hence elites [95].

The data analysis of the news dataset supports a strong correlation between overrepresentation of managerial and economic frames and justice. The scholarship that has emerged to support the rationale for why these correlations exist rest in social command frames and social contest frames. Although this research utilized standard communication frames, a generalization of these frames is encapsulated under three categories: social command frames, social contest frames, and social body frames.

Social command frames are viewed as showing dominance or submission, strengths and weaknesses, or independence or dependence [95]. Given this insight, frames such as managerial and economic tend to fall into the social command frames because managerial frame choice is associated with authority dominance. When online news reporters use managerial frames in conjunction with economic frames, an embeddedness of power exists in operation where control and coercion underpin the amplification of subjective texts.

In contrast, conflict frame choice aligns with social contest frames. In this sense, frames are devised instruments of competition and games (or winners and losers). The conflict frames thrive under the notion that opposition is normal and necessary to get anything done. Therefore, online news articles embodying this frame tend to convey ideas about fighting, war, rewarding truth, excellence, innovation, and productivity [94].

The human interest and scientific frames fall under social body frames, which are best characterized as humans being collectively intertwined and connected as one. Commonly used language under this banner is about the common good or well-being of all, love thy neighbor as yourself, or namaste—I bow to the divine in you.

The strategic use of managerial, economic, or conflict frames as expressivity in language embedded within online news, particularly when reporting on critical issues that impact society, has grave costs and consequences, which are far reaching. In the end, society's perceptions, beliefs, and attitudes become polarized or conflicted about critical issues, even in the same domain, rather than society coming to a common consensus. According to Lakoff [93], deep frame structures—where moral foundations exist—must

be in operation in the mind during the production of language for justice about critical issues to prevail.

These emerging insights are justification for using justice in relation to frame identify as forces for understanding shifts in amplification intensity that lead to programmable insights about online news sources reporting on critical issues.

Fig. 19 graphically displays the possibility space of cusp model, an extension of the cusp catastrophe base model.

Let $M = \{Managerial, Economic, Conflict\}$ and $S = \{Human\ Interest, Science\}$ be the aggregate count of the sets of frame choices obtained from the edge connection. S_g is assigned the frame identity corresponding to its F_i property at time t.

Definition: Amplification is represented by $\boldsymbol{z}$, a projection on the behavior surface for predicting the migration path that shows the gradual shift in amplification intensity of the online news source when shaping the frame narrative about the issue of water insecurity. This factor is a measure of the strength of amplification—that is, "weak amplification" is considered safe and is represented by *S(ecure)*, and "strong amplification" is considered insecure or harmful and is represented by *I(nsecure)*.

Definition: Asymmetry control (or normal) factor is represented by $\boldsymbol{x}$. This factor receives the label "Justice," denoted on a scale that ranges from $J(ust)$ to $U(njust)$. Justice is a measure of the news source's perception of fairness of water consequences.

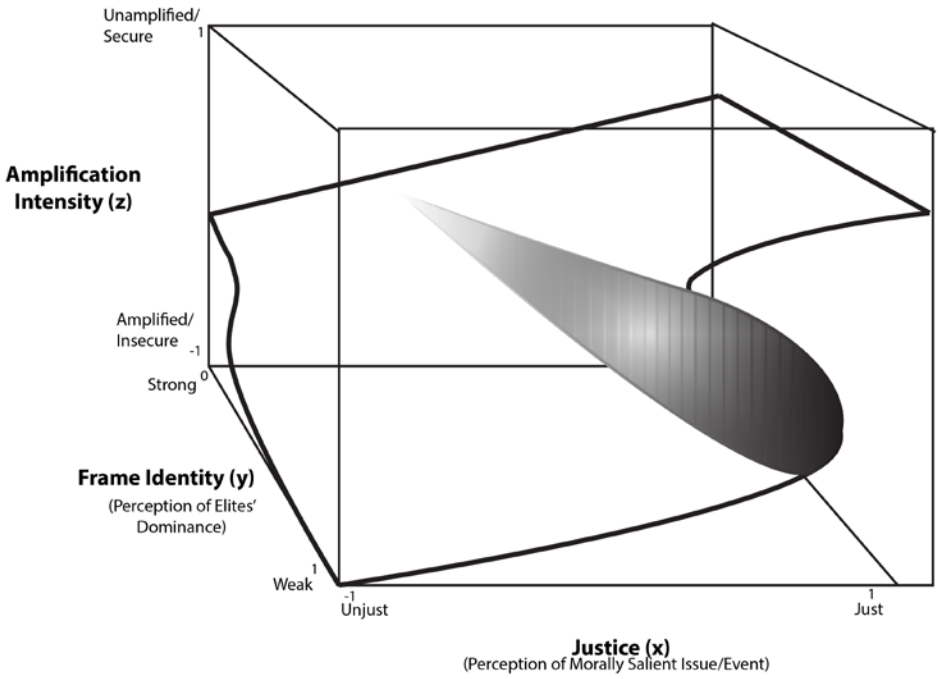


Fig.19. The image depicts the cusp catastrophe model for measuring the shift in news sources amplification intensity.

Definition*:* Splitting factor or bifurcation factor is represented by **y**. This factor receives the label "Frame Identity," denoted as $S(trong)$ or $W(eak)$. This is the perceived dominance by elites when reporting the news.

Here, I argue that frame identity and the perception of justice are key forces for predicting the amplification intensity of news sources when producing articles. Frame identity is one form of amplification; it is shaped by strategic devices for presenting prominent aspects and perspectives about an issue using a strong slant for the purpose of conveying latent meanings about an issue [30]. The position supported in this study is that justice is of equal importance, as it captures the perceived sense of fairness about an issue.

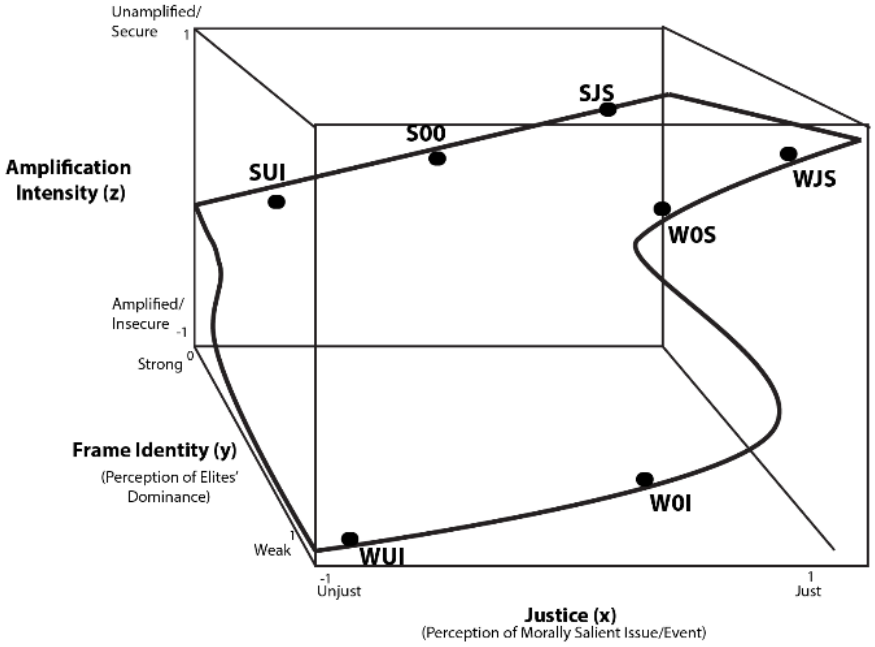


Fig. 20. This is a depiction of the state in the possibility space (surface space) for showing correlation of Frame Identity and Justice forces that may cause gradual shifts in amplification of framing of the issue.

Fig. 20 provides a depiction of the news source amplification intensity in accordance with the CCM for predicting gradual shifts of views and showing the possibility space. This allows one to visualize discrete points in time over the possibility space. A discretized approach to opinion research has been conducted in scholarly research [40] and applications of the CCM [41]–[42].

E. Possibility Space Transitions

Nonlinear relationships represent a variety of possibilities and pathways of which the linear case is a limiting one. I express the possibility space for the research study using three variables as shown in Table 1.

Table 1
Possibility Space Index. (a) Frame Identity (Splitting Factor). (b) Justice (Normal Factor). (c) Amplification Intensity

(a)	Index-1	Description	Values
	S	S(trong)	0
	W	W(eak)	1
(b)	Index-2	Description	Values
	J	J(ust)	1
	U	U(njust)	−1
	0	N(eutral)	0
(c)	Index-3	Description	Values
	S	S(ecure)	1
	I	I(nsecure)	−1
	0	N(eutral)	0

Table 1. Explanation of index meanings. (a) Index 1 represents the control variable "frame identity" (splitting factor), which is the perceived dominance by elites when reporting the news. (b) Index 2 represents the sense of "justice" (normal factor) in the handling of the issue. (c) Index 3 represents the dependent variable "amplification intensity," which has the effect of signaling a secure (or safe) or insecure (or harmful) issue.

Index-1, Index-2, and Index-3 are combined to denote state transitions for updating shrinkage and growth. For instance, SJS denotes strong frame identity (S), a sense of justness (J), and a sense of security or being safe (S).

News sources' selection and influence occur through signaling news message amplification intensity. Table 2 indicates where news sources change amplification intensity based on interactions. Each column shows location of growth in the number of recipients switching from one news source to another, as a result of observed amplification intensity. The shrinkage and growth at a single site are proportional to the number of recipients and the number of senders signaling. Table 2 illustrates the shrinkages and growth that may occur over time.

Table 2
Possibility State Space for Migrating the Network
Signaling Content (State of News Sender)

Recipient	**SUI**	**S00**	**SJS**	**WUI**	**W0I**	**W0S**	**WJS**
SUI			S00			WUI	W0I
S00	SUI		SJS				
SJS	S00			W0S	WJS		
WUI	SUI		W0I				W0I
W01	S00						
W0S			S00				
WJS	W0S		W0S	W0S			

VII. Visual Language Design

Traces of visualization for communicating ideas predates to thousands of years. Recently, attention has turned to the usage of visualization for audience influence and persuasion. Most importantly, with the influx of big data on the Internet, visualization has become the premier tool for exploratory data analysis that enables the formation of hypotheses about a network.

Visualization provides special usability benefits, particularly when the goal is to synthesize vast amounts of data over time. Visualization can provide insights into the spatial and temporal informational flows and dynamics; they allow for inference and discovery. The properties of visual, spatial, and time are essential to building human intelligence. This research introduces the double subjectivity social influence model for exploring the interplay of the dynamic system about critical issues—in this case, water insecurity.

In this research, a visual language that comprises alphabets for expressing nodes and grammar for expressing relational edge connections are presented. It is used for communicating the cumulative effects of double subjectivity for understanding the spatiotemporal dynamics underlying the network of the collective influence of framing in online news and as a factor in the production of language in text narratives.

The overarching research inquiry may be explored and is expressed through the construction of an online news frame issues network. Table 3 includes the visual language for expressing and exploring the double subjectivity social influence model

using the issues network for understanding the network structure and underpinning dynamic forces.

Table 3
Visual Language Notations, Symbols, and Definitions

Notation	Symbol	Description
S_g		Thick blue circles around the notation S_g denote online news the sources node that produced the article and propagates a central organizing idea (i.e., framing).
A_h		Thin orange circles around the notation A_h denote an online news articles node.
E_1		Solid black lines correspond to edge connections among neighboring articles covering a similar issue.
E_2		Solid lines with colors {red, green, purple} correspond to edge connections of news sources who produce articles covering a similar issue within domains.
E_3		Dotted lines with colors {red, blue, green, purple, orange} correspond to news frame choice (or strategy). For the issue of water insecurity, standard frame types are used (An & Gower, 2009)—in particular, human interest, conflict, economic, managerial, and science.

The data analyst is generally interested in exploring the network structure. Given behavior of news sources when reporting on critical issues, like water insecurity, is complex; the interactions are not static, even when measured for very short intervals. The double subjectivity social influence model makes visible dynamics that may be easily missed in big data platforms, such as the Internet. The instruments identified in Table 4

may be used as probes for gaining programmable insights into the network and the forces at work.

TABLE 4
Instruments for Measuring

Notation	Symbol	Description
G_t		Graph composed of nodes and edges that arrived before time t
S		Total number of online news source nodes in a graph
A		Total number of online news articles nodes in a graph
E_a		Total number of article-to-article edges in a graph
E_s		Total number of article-to-source edges in a graph
E_f		Total number of source edges employing the same frame in a graph
$S_g(t)$		Number of online news source nodes in a graph at time t
$A_h\ (t)$		Number of online news article nodes in a graph at time t
$A\ (e)$		Number of article edges in a graph at time t
$S\ (e)$		Number of source edges in a graph at time t
$F\ (e)$		Number of frame edges in a graph at time t
b		Amplification Intensity Constant
$f(b)$		Amplification Intensity Function

This research offers a method for time stepwise transitions when navigating the dynamic online news issues network, as shown in Fig. 20. The double subjectivity social influence model provides a mechanism for control over belief learning using the values

of α_g, while β_g offers news sources inside signaling mechanism by leveraging historical knowledge about neighboring nodes.

In Fig. 21, an illustration is given to show scenarios where the beliefs of news sources converge to a state of consensus, as shown in Fig. 21(a), and conflict, as shown in Fig. 21(b). For this illustration, one may use a Euclidean distance measure for showing the degree of change among news sources.

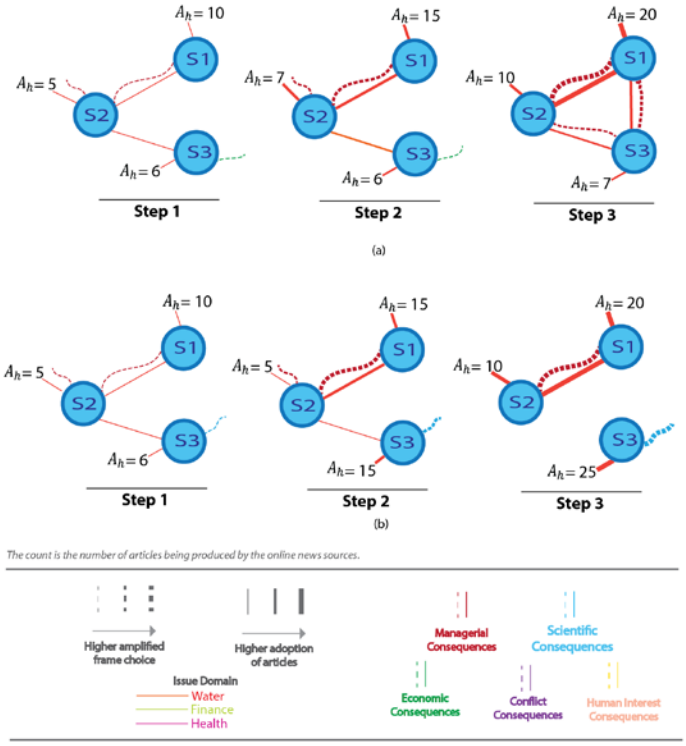


Fig. 21. Visual representation of the model. This is an illustrative scenario of news source interactions where double subjectivity has spatiotemporal dynamics and social influence with causal effect. Diagram (a) shows interactions that result in consensus whereas (b) shows the formation of an alternative path that may be expressed as conflict in beliefs or attitudes about the issue of water. This depiction shows the five standard communications frames used in this research and three domains where vast amounts of narrative unstructured text exist online.

VIII. Observations of the Models

From the search of articles for keywords, 55 000 news articles were collected and stored in a database for preprocessing. After preprocessing, 30 000 articles were deemed relevant for the dataset; 280 news sources were identified, using strategic framing, on the basis of their potential to influence public perceptions of the issues of water insecurity.

In this paper, models of cultural dynamics are integrated with the CCM to explore double subjectivity through quantifying the amplification intensity. The analysis shows that this measure has potential to change the network structure, as prominent mass count is not the only law of attraction at work.

The power of double subjectivity is in the deep learning of the entire network; both in spatial and temporal terms. This research may offer insights into framing amplification that are overlooked in online narrative text over time. Two ways for gaining deeper knowledge are offered in this context: 1) a visualization for exploring how news framing promotes prominence in a network structure and 2) a model whereby forces may be examined when quantifying amplification and the migration path of amplification formation.

For instance, if one desired to switch to the financial domain and, specifically, learning about prominent personas in this domain, such a tool as discussed in this research may be helpful. Consider the case of Bernard Madoff, a former stockbroker, investment advisor, and financier. Bernard Madoff's Ponzi scheme exemplifies a case in which there may have been "red flags" that were overlooked in letters and outlook sections of the financial annual reports produced by Madoff. Madoff held prominence as an elite in the financial markets network, which was a driving force for shaping public

opinion about his decisions. Thus, the spread of his decisions influenced others in the financial markets. Fig. 22 illustrates an example of the outcome of the double subjectivity social influence model as characterized by the sociotemporal dynamics graph and its visualization; an indirect graph consisting of source and author nodes is visualized over three time steps. The number of nodes remains consistent across each time step. Here, the mass experience grows (gains) or shrinks (loss) because of gradual shifts in frame amplification—in this case, the forces are frame identify and justice. This has the cumulative effect of double subjectivity.

Although the double subjectivity social influence model is presented to scale to full dataset capacity, a small subset of the network is used for this preliminary evaluation. The model allows for news sources that generate articles about water consequences to interact freely with other news sources. However, because news sources receive signals

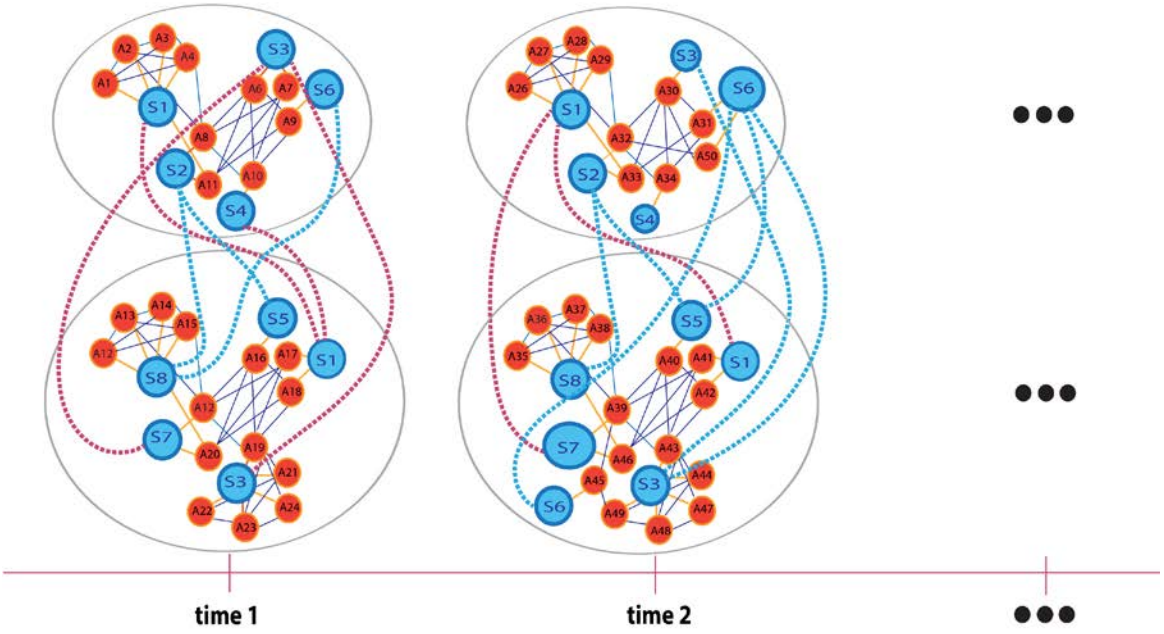


Fig. 22. Transition graph of the social influence and migration path for showing the mass growth and shrinkage process.

that provide a relative baseline—leveraging other news sources' opinions, interpretations, and perceptions—it is most likely that a news source will choose interactions with similar sources. A chance exists for a news source to interact with another news source based on an increase in its amplification intensity although no connection exists. However, the chance is higher that news sources will interact only with connected sources; thus, breaking out of the cluster is caused by catastrophic shifts in amplification.

With the amplification intensity function, a variant of the CCM shows possibility space in behavior when making decisions about trust and distrust. Any space on the surface response represents the state of the observed news source. As the news source perceives that the issue will not lead to fair water consequences, the cusp triggers them to start attributing blame toward the population causing harm. The contextualization of this case with water insecurity shows that a news source with ties to nodes that employ the frame choice of perceived unfairness (unjust) will notice that the sources more aligned with elites are becoming increasingly bullish on the water issue. These news sources who initially framed the issue as unjust attract other news sources (i.e., homophily) because their narrative is spreading. When feelings of unfairness intersect with the dominance of the elites, the news sources start scanning for those to whom to attribute blame, thereby entering a space of distrust and strong amplification. This kind of model is associated with hysteresis, making it hard to shift between surface response planes, as the migration path will be different.

IX. Summary of Claims

This dissertation presents 1) the first formal computer science definition for framing, 2) an approach for the discovery of five distinct patterns (referred to as signatures) that characterize prominent frames, and 3) a process that advances machine learning for document classification that takes advantage of content analysis on a small number of documents for scaling up to meet the demands of large datasets. Preliminary experiments suggest the use of NMF combined with TF-IDF are promising for discovering frames through the process of revealing latent relationships found in online news articles. Though research by computer scientists is underway in using communication frames to explore news discourse [52]–[54], those studies fail to build upon a basic formal definition of framing that is understandable by the machine for learning and in the context of computer science. Thus, this research fills an important gap in the computer science literature by providing a formal definition and process for discovering frames.

In this dissertation, a new concept was presented that allows for an adjustment of the absolute baseline of a news source (or agent) that takes into account one′s amplification for expressing double subjectivity through news frames. Further, offered in this paper is 1) the first known formal news frame issues network in computer science, 2) a model for learning the migration paths and patterns about issues, and 3) the first known formulation of an amplification intensity function using the CCM for showing the possibility space and gradual shifts in views that are sensitive to news framing amplification. Preliminary experiments suggest that the integration of cultural dynamic

models with a CCM is promising for exploring double subjectivity for revealing latent relationships found in online news articles.

This research extends some capabilities of the formal computational world into the informal, subjective cognitive world by allowing users to formally process and automate some subjective concepts using software. The methodologies and model may be utilized if one were to pursue the opposite direction by adopting some concepts from the subjective world into the formal domain. A realization of this extension is through the usage of visualization for simplifying the human interactions. This research offers a methodology for extracting the subjectivity from unstructured narrative text using formal methods. Once processed through the double subjectivity tool, the output is objective and formal, which, in turn, is displayed in the humanized output using graphic visualization for informal representation.

Future work involves improving the double subjectivity social influence model that considers multiple variables and forces that expose a greater possibility space for deep learning. Further testing of other hypotheses and unfolding the potential function to explore conditions of convergence may be valuable. In addition, future work will give treatment to other hypotheses associated with amplification that leads to trust or distrust, because of perceived positive consequences about the critical issue under study.

These techniques and methods may be adapted to compare classification for discovering dominant frames using multimedia big data classification, particularly the use of a classifier ensemble framework.

X. Personas for Double Subjectivity Framework

The double subjectivity social influence model may be tailored for specific uses. Here, a brief discussion is given for the motivations for using the model discussed in this research for 1) discovering dominant frames in online news, 2) gaining programmable insights about alternative pathways for the production of language for advancing cognitive computation, 3) unveiling hidden meanings in narrative text, 4) and gaining programmable insights by exploring spatiotemporal dynamics when narrative text intensifies that may have the effect of slanting the truth and ultimately shifting attitudes, beliefs, and values.

Marketing. Marketing giant Adobe, a service provider for advertising agencies, is interested in ensuring their customers receive a return on their investment (ROI): a measure of reachability and conversion rate. They support the construction of messaging and strategy for advertisers to promote the sale of products, goods, and services. To date, Adobe does not have tools for ensuring this level of service quality. This research offers a tool for Adobe to explore in real time a visualization of the social influence and dynamics of their marketing messaging in online news.

Financial Markets. The concept of double subjectivity has considerable currency to stock market investors as they have an interest in understanding their risk exposure in advance of crises or stock crashes. The double subjectivity framework may be useful for learning the embedded internal structure of unstructured narrative financial text—earning call transcripts, annual reports, letters to shareholders, and disclosure statements—that contain valuable embedded information cues and frame amplification about the company

performance. Currently, this body of narrative text remains untapped due in large part to human limitations of readability. The human ability is unprecedented, but unscalable. Moreover, the double subjectivity framework visualization will expose the migration path of financial narrative text, the amplification intensity, and the intersection of forces where shifts in beliefs are realized.

Activism and Social Movements. "Activist" is the persona who could benefit from the double subjectivity framework. The intent of social movements is to place demands for change on organizations or institutions through collective action. There is a tremendous amount of implicit frame amplification flowing in both online news and in social media content when social movements erupt. The double subjectivity framework can provide important insight into the framing effects of a social movement issue that may influence the choices people make and the consequential actions taken. Most importantly, this tool will expose in real time the collective and cumulative effects of amplification framing that flow in online text for causing change in public perceptions about the issue; the targeted organizations or institutions can gauge how the activism influences public policy.

REFERENCES

[1] P. Clifford and A. Sudbury, A Model for Spatial Conflict, Biometrika, 60(3):581-588, 1973.

[2] Eli Ben-Naim, Paul Krapivsky, and Sidney Redner. Bifurcations and patterns in compromise processes. Physica D, 183, 2003.

[3] C. Castellano, S. Fortunato, and V. Loreto, Statistical physics of social dynamics. Reviews of Modern Physics, 81:591–646, 2009.

[4] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, Mixing beliefs among interacting agents, Advances in Complex Systems, 3:87–98, 2000.

[5] R. Hegselmann and U. Krause, Opinion dynamics and bounded confidence: Models, analysis and simulation, Journal of Artificial Societies and Social Simulation, 5(3), 2002.

[6] B. Latane′. The psychology of social impact. American psychologist, 36(4):343, 1981.

[7] R Axelrod. The dissemination of culture - A model with local convergence and global polarization. Journal of Confict Resolution, 41(2):203{226, APR 1997. ISSN 0022-0027. doi: f10.1177/0022002797041002001g.

[8] K. C. Chen, M. Chiang, H. V. Poor, “From technological networks to social networks,” published in IEEE J. Sel. Areas Commun.., 2013.

[9] Y. F. Chou, R. G. Cheng, and H. H. Huang, “Performance evaluation of referral selection algorithms in human centric communication networks,”in Proc. 2012 International Symposium on Wireless Personal Multimedia Communications, pp. 638–642, 2012.

[10] M. McPherson, L. Smith-Lovin, and J. M. Cook, Birds of a feather: Homophily in social networks, Annual Review of Sociology, 27:415–444, 2001.

[11] S. Mullainathan and A. Shleifer, “The Market for News.” American Economic Review 95 (1): 1031-1053, 2005.

[12] T. A. DiPrete and G. M. Eirich, Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments, Annual Review of Sociology, Vol. 32: 271-297, Volume publication date August 2006.

[13] L. R. Rabinerand B. H. Juang,”An introduction to hidden Markovmodels,” IEEE ASSP Mag., vol. 3, no. l, pp. 4-16, 1986.

[14] P. Raghavan, S. Amer-Yahia and L. Gravano eds., “Structure in Text: Extraction and Exploitation”, In. Proceeding of the 7th international Workshop on the Web and Databases(WebDB), ACM SIGMOD/PODS 2004, ACM Press, Vol 67, 2004.

[15] H. P. Luhn. “The Automatic Creation of Literature Abstracts,” IBM Journal of Research and Development, 2, No. 2, 159, April 1958.

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

[17] R. Steinheiser, and C. Clifton, "Data Mining on Text." 2012 IEEE 36th Annual Computer Software and Applications Conference IEEE Computer Society, p. 630, 1998.

[18] A. Tan, N. Zhong and L. Zhou, "Text Mining: The state of the art and the challenges." Proceedings of the PAKDD Workshop on Knowledge Discovery and Data Mining, 65—70, Beijing, China 2000.

[19] S. Tan, X. Cheng, B. Wang, H. Xu, M. M. Ghanem, Y. Guo, “Using dragpushing to refine centroid text classifiers”, In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '05). ACM, New York, NY, USA, 653-654. 2005.

[20] M. Tsytsarau & T. Palpanas, Survey on mining subjective data on the web, Data Mining and Knowledge Discovery, 24(3), 478-514, 2012.

[21] H. Karanikas, C. Tjortjis, and B. Theodoulidis. "An Approach to Text Mining using Information Extraction", Proc. Workshop Knowledge Management Theory Applications (KMTA 00), 2000.

[22] Q. Hu, D. Yu, Y. Duan, & W. Bao, A novel weighting formula and feature selection for text classification based on rough set theory, Natural Language Processing and Knowledge Engineering, 2003, Proceedings. 2003 International Conference on. IEEE, 2003.

[23] B. Liu, Sentiment analysis and opinion mining, Synthesis lectures on human language technologies, 5(1), 1-167, 2012.

[24] D. A. Scheufele, “Framing as a theory of media effects”, The Journal of Communication, 49(1), 103-122, 1999.

[25] S. Dumais, G.W. Furnas, T.K. Landauer, S. Deerwester, Using latent semantic analysis to improve information retrieval. In Proceedings of CHI’88: Conference on Human Factors in Computing, New York: ACM, 281-285, 1988.

[26] C. Fellbaum, “A Semantic Network of English: The Mother of all WordNets,” in: Computers and the Humanities 32: 209-220. 1998.

[27] D. Bindela, J. Kleinberg, & S. Orenb, How bad is forming your own opinion?, Games and Economic Behavior, 92, 248–265, 2015.

[28] D. Kempe, J. Kleinberg, and ´ E. Tardos, “Maximizing the spread of influence through a social network,” in Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 137–146, 2003.

[29] M. Lombard, J. Snyder-Duch, C. C. Bracken, Content analysis in mass communication: Assessment and reporting of intercoder reliability. Human Communication Research, 28, 587–604, 2002.

[30] R. M. Entman, “Framing: Towards clarification of a fractured paradigm”, Journal of Communication, 43(4), 51-58, 1993.

[31] G.F. Bishop, R.W. Oldendick, A.J.Tuchfarber,. Effects of presenting one versus two sides of an issue in survey questions. The Public Opinion Quarterly 46, 69–85, 1982.

[32] J.A., Krosnick, Survey research. Annu. Rev. Psychol. 50, 537–567, 1999.

[33] K. C. Chen, M. Chiang, H. V. Poor, “From technological networks to social networks,” accepted and to be published in IEEE J. Sel. Areas Commun., 2013.

[34] Y. F. Chou, R. G. Cheng, and H. H. Huang, “Performance evaluation of referral selection algorithms in human centric communication networks,”in Proc. 2012 International Symposium on Wireless Personal Multimedia Communications, pp. 638–642, 2012.

[35] H. Liu, P. Singh, Focusing on conceptnet’s natural language knowledge representation. In Proc. Of the 8th Intl Conf. on Knowledge-Based Intelligent Information and Engineering Syst, 2004.

[36] R. K.Garrett, B. E. Weeks, R. L. Neo, Driving a Wedge Between Evidence and Beliefs: How Online Ideological News Exposure Promotes Political Misperceptions, Journal of Computer-Mediated Communication, 4 August 2016.

[37] I. Stewart, Does God Play Dice?: The Mathematics of Chaos, Basil Blackwell (Oxford), pp 84–94, 1989.

[38] R. Thom. Structural Stability and Morphogenesis, 1972. D. H. Fowler (trans.). Reading, MA: W.A. Benjamin.

[39] L. Skitka & E. Sargis, The internet as psychological laboratory. Annu Rev Psychol 57: 529–555, 2006.

[40] P. Sobkowicz, Discrete model of opinion changes using knowledge and emotions as control variables. PloS one, 7(9), e44489, 2012.

[41] T. F. Homer-Dixon, Environment, Scarcity, and Violence. Princeton University Press, 1999.

[42] Homer-Dixon, T. F., The upside of down: Catastrophe, creativity, and the renewal of civilization. Washington: Island Press, 2006.

[43] Newman, M. E. J., Networks an Introduction. Oxford University Press, 2011.

[44] L. H. Cheeks, T. L. Stepien, D. M. Wald, and A. Gaffar, Discovering news frames: An approach for exploring text, content, and concepts in online news sources, International Journal of Multimedia Data Engineering and Management (IJMDEM), 7, pp. 45–62, 2016.

[45] D. Chong & J. Druckman, Framing public opinion in competitive democracies. American Political Science Review, 101(4), 637-655, 2007a.

[46] D. Chong & J. Druckman, A theory of framing and opinion formation in competitive and elite environments. Journal of Communication, 57, 99-118, 2007b.

[47] B. K. Jaworski, Media Framing Of U.S. Health Care Reform: A New Era Or Reinforcing Dominant Ideologies Of Health And The Health Care System?, 2012.

[49] K. Krippendorff, Content analysis: An introduction to its methodology, Beverly Hills, CA: Sage, 1980.

[50] L. H. Cheeks & A. Gaffar, A Social Influence Model for Exploring Double Subjectivity Through News Frames in Online News Sources, IEEE Intelligent Systems Conference (IntelliSys), 2017.

[51] G. Lakoff, Why it Matters How We Frame the Environment, Environmental Communication Vol. 4, No. 1, pp. 7081, March 2010.

[52] Oracle Corporation, WWW, www.oracle.com, 2008.

[53] D. P. Cartwright, “Analysis of qualitative material”, In L. Festinger & D. Katz (Eds.), Research Methods in the Behavioral Sciences (pp. 421-470). New York: Dryden, 1953.

[54] J. Tankard, The empirical approach to the study of media framing. In S. Reese, O. Gandy, & A. Grant (Eds.), Framing public life (pp. 95–106), 2001, Mahwah, NJ: Erlbaum.

[55] W. J. Severin, J. W. Tankard, Communication theories: origins, methods, and uses in the mass media, 4th Edition, Longman Publishing Group, New York, 1997.

[56] S. K. An, K. K. Gower, “How do the news media frame crises? A content analysis of crisis news coverage”, Public Relations Review 35, 107–112, 2009.

[57] B. G. Glaser, A. L. Strauss, The Discovery of Grounded Theory; Strategies for Qualitative Research, Chicago: Aldine Pub. Co., 1967.

[58] O. R. Holsti, Content Analysis for the Social Sciences and Humanities, Reading, MA: Addison-Wesley, 1969.

[59] J. N. Druckman, T. Bolsen. Framing, motivated reasoning, and opinions about emergent technologies. Journal of Communication, 2002.

[60] D. A. Scheufele, D. Tewksbury, Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models. Journal of Communication 57 (1):9-20, 2007.

[61] T. E. Patterson, Pre-Primary News Coverage of the 2016 Presidential Race: Trump’s Rise, Sanders’ Emergence, Clinton’s Struggle, Shorenstein Center on Media, Politics and Public Policy in conjunction with Media Tenor, 2016.

[62] Quoted in T. E. Patterson, “More Style Than Substance,” Political Communication and Persuasion 8, p.157, 1991.

[63] In presidential contest, voters say ‘basic facts,’ not just policies, are in dispute, Pew Research Centre, October 2016.

[64] F. Manjoo, “How the Internet Is Loosening Our Grip on the Truth,” New York Times, Nov. 2, 2016. http://www.nytimes.com/2016/11/03/technology/how-the-internet-is-loosening-our-grip-on-the-truth.html

[65] P. Murphy, “Affiliation bias and expert disagreement in framing the nicotine addiction debate”, Science, Technology, & Human Values, 26, 278–299, 2001.

[66] D. Shah, M. D. Watts, D. Domke, D. P. Fan, “News framing and cueing of issue regimes: Explaining Clinton’s public approval in spite of scandal”. Public Opinion Quarterly, 66(3), 339–370, 2002.

[67] P. Baxendale, Machine-made index for technical literature - an experiment, IBM Journal of Research Development, 2(4):354–361, 1958.

[68] M. E. Maron, J. L. Kuhns, "On Relevance, Probabilistic Indexing and Information Retrieval", Journal of the ACM, 216-244, 1960.

[69] H.P. Edmundson, New methods in automatic extracting, Journal of the ACM, 16(2), 264–285, 1969.

[70] K. S. Jones, “A statistical interpretation of term specificity and its application in retrieval”, Journal of Documentation, Vol. 28, pp. 11–21, 1972.

[71] Q. Hu, D. Yu, Y. Duan and W. Bao, "A novel weighting formula and feature selection for text classification based on rough set theory," International Conference on Natural Language Processing and Knowledge Engineering, 2003. Proceedings. 2003, Beijing, China, pp. 638-645, 2003.

[72] A. Hotho, S. Staab and G. Stumme, "Ontologies improve text document clustering," Third IEEE International Conference on Data Mining, pp. 541-544, 2003.

[73] C. Luo, Y. Li, S. M. Chung, "Text document clustering based on neighbors" Data & Knowledge Engineering, 68.11:1271-1288, 2009.

[74] C. Jiang, F. Coenen, R. Sanderson, M. Zito, “Text Classification Using Graph Mining-Based Feature Extraction”, Knowledge-Based Systems, vol. 23, no. 4, pp. 302-308, 2010.

[75] B. Berendt, A. Hotho, G. Stumme, "Towards semantic web mining", The Semantic Web—ISWC 2002, Springer Berlin Heidelberg, pp. 264-278. 2002.

[76] K. Dave, L. Steve, D. M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews", Proceedings of the 12th international conference on World Wide Web. ACM, 2003.

[78] K. H. Krippendorff, Content Analysis - 3rd Edition: An Introduction to Its Methodology, Thousand Oaks: SAGE Publications, Inc, 2013. Print.

[79] D. Riffe, S. Lacy, F. G. Fico, Analyzing Media Messages: Using Quantitative Content Analysis in Research, New York, NY: Routledge, 2005.

[80] D. Riffe, S. Lacy, F. G. Fico, Analyzing Media Messages, 3rd Edition, 2013,

[81] M. A. Hearst, “Text data mining: Issues, techniques, and the relationship to information access”, Presentation notes for UW/MS workshop on data mining, July 1997.

[82] P. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining, Boston : Pearson Addison Wesley, c2006.

[83] D. Kempe, J. Kleinberg, S. Oren. and A. Slivkins, Selection and influence in cultural dynamics. Network Science, 4(01), pp.1-27, 2016.

[84] D. Blei, A. Ng, and M. Jordan, “Latent Dirichlet allocation,” Journal of Machine Learning Research, vol. 3, pp. 993–1022, 2003.

[85] A. Stavrianou, P. Andritsos, N. Nicoloyannis, “Overview and Semantic Issues of Text Mining”, Special Interest Group Management of Data (SIGMOD) Record, (Vol. 36, No. 3), September 2007.

[86] P. Murphy, “Affiliation bias and expert disagreement in framing the nicotine addiction debate”, Science, Technology, & Human Values, 26, 278–299, 2001.

[87] W.J. Severin, J.W. Tankard, Communication theories: origins, methods, and uses in the mass media, 4th Edition, Longman Publishing Group, New York, 1997.

[88] W. A. Gamson, A. Modigliani, The Changing Culture Of Affirmative Action. Research in Political Sociology, 3, 137–177, 1987.

[89] A. Karlberg, News and Conflict, How adversarial News Frames Limit Public understanding of Environmental Issues, Alternative Journals, 1997.

[90] D. P. Bertsekas, Nonlinear Programming, Athena Scientific, Belmont, MA 02178-9998, second edition, 1999.

[91] P. O. Hoyer, “Non-negative Matrix Factorization with Sparseness Constraints”, Journal of Machine Learning Research 5, 1457-1469, 2004.

[92] H. Liu, Z. Wu, “Non-negative matrix factorization with constraints,” in Twenty-Fourth AAAI Conference on Artificial Intelligence, 2010.

[93] G. Lakoff, Whose Freedom? The Battle Over America′s Most Important Idea, New York: Farra, Straus & Giroux, 2006a.

[94] M. Karl, Constructive Resilience: The Bah′a′I Response to Oppression, Peace and Change, 35, no. 2:222-257, 2010.

———, ″Western Liberal Democracy as New World Order?″ In The Bahá′í World: 2005–2006, edited by Robert Weinberg, 133–156. Haifa, Israel: World Center Publications, 2007.

———, ″The Power of Discourse and the Discourse of Power: Pursuing Peace through Discourse Intervention.″ International Journal of Peace Studies 10, no. 1: 1–25, 2005.

[95] M. Karlberg, Forming a Culture of Peace, Equity, and Justice (ed.), Palgrave MacMillan, 2012.

[96] Y. H. Chen, L. Chen, X. Q. Wang, and G. Chen, Critical buckling load calculation of piles based on cusp catastrophe theory, Mar Georesour Geotechnol, 33: 222−228, 2013.

[97] D. Jiang, G. Hao, L. Huang, D. Zhang, Use of Cusp Catastrophe for Risk Analysis of Navigational Environment: A Case Study of Three Gorges Reservoir Area, 2016.