

From social media to expert reports:
Automatically validating and extending complex
conceptual models using machine learning approaches

by

Mannila Sandhu
Lakehead University

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTERS

in the Department of Computer Science



All rights reserved. This thesis may not be reproduced in whole or in part, by
photocopying or other means, without the permission of the author.

From social media to expert reports:
Automatically validating and extending complex
conceptual models using machine learning approaches

by

Mannila Sandhu
Lakehead University

Supervisory Committee

Dr. Vijay Mago, Supervisor
(Department of Computer Science, Lakehead University, Canada)

Dr. Philippe J. Giabbanelli, Co-Supervisor
(Department of Computer Science, Miami University, USA)

Dr. Zubair Fadlullah, Departmental Member
(Department of Computer Science, Lakehead University, Canada)

Dr. Abdulsalam Yassine, External Member
(Department of Software Engineering, Lakehead University, Canada)

ABSTRACT

Given the importance of developing accurate models of any complex system, the modeling process often seeks to be comprehensive by including experts and community members. While many qualitative modeling processes can produce models in the form of maps (e.g., cognitive/concept mapping, causal loop diagrams), they are generally conducted with a facilitator. The limited capacity of the facilitators limits the number of participants. The need to be either physically present (for face-to-face sessions) or at least in a compatible time zone (for phone interviews) also limits the geographical diversity of participants. In addition, participants may not openly express their beliefs (e.g., weight discrimination, political views) when perceiving that they may not be well received by a facilitator or others in the room. In contrast, the naturally occurring exchange of perspectives on social media provides an unobtrusive approach to collecting beliefs on causes and consequences of such complex systems. Mining social media also supports a scalable approach and a geographically diverse sample. While obtaining a conceptual model via social media can inform policymakers about popular support for possible policies, the model may stand in stark contrast with an expert-based model. Identifying and reconciling these differences is an important step to integrate social computing with policy making.

The pipeline to automatically validate large conceptual models, here of obesity and politics using large text data-set (academic reports or social media like Twitter) comprise technical innovation of applying machine learning approaches. This is achieved by generating relevant keywords using wordnet interface from NLTK, articulating topic modelling using gensim LDA model, entity recognition using Google Cloud Natural language processing API and categorizing themes by count vectorizer and tf-idf transformer using scikit-learn library. Once the pipeline validates the model, it is further suggested for extension by mining literature or Twitter conversations and using Granger causality tests on the time series gained from respective sources of data. Later we realize the impact of the shift in public opinion on Twitter, which can alter the results of validation and extension of conceptual models while using our computational methods. So we finally compare the sentiment analysis and sarcasm detection results on these conceptual models. Analyzing these results we discuss whether the confirmed and extended associations in our conceptual model are an artifact of our method or an accurate reflection of events related to that complex conceptual model. The combination of these machine

learning approaches will help us automatically confirm and extend complex conceptual models with less hassle of money, time and resources. It can be used for automatically formulating public policies which are created in response to issues brought before decision makers, instead we create them using issues discussed everyday on social media platform.

Contents

Supervisory Committee	ii
Abstract	iii
Table of Contents	v
List of Tables	viii
List of Figures	ix
Acknowledgements	xi
Dedication	xii
Thesis Structure	xiii
1 Introduction	1
1.1 Background	1
1.2 Objectives	4
1.3 Main Contributions	5
2 From social media to expert reports: the impact of source selection on automatically validating complex conceptual models of obesity	6
2.1 Abstract	7
2.2 Introduction	7
2.3 Background	11
2.3.1 Social web mining for health	11
2.3.2 Conceptual models in obesity research	13
2.4 Validating a conceptual model from text	14
2.5 Comparing conceptual models from Twitter and expert reports	17

2.5.1	Data-sets and pre-processing	17
2.5.2	Validating the model for each data-set	19
2.6	Discussion	22
2.7	Conclusion	24
3	Validating conceptual models by mining Twitter data	25
3.1	Abstract	26
3.2	Introduction	26
3.3	Methods	29
3.3.1	Overview	29
3.3.2	Representing nodes and edges of user concept map	31
3.3.3	Mapping derivationally related form onto concept nodes in lemma form	33
3.3.4	Mapping relevant tweets onto concept nodes	34
3.3.5	Extracting themes from mapped relevant tweets	35
3.3.6	Validating edges of user concept map	36
3.4	Experiment	36
3.4.1	Experimental set-up	36
3.5	Implementation	40
3.6	Discussion	43
3.7	Conclusion	45
4	Extending conceptual models using Granger causality by mining lit- erature and Twitter conversations	48
4.1	Abstract	48
4.2	Introduction	49
4.3	Methodology	50
4.3.1	Data Collection	50
4.3.2	Data preprocessing	53
4.3.3	Test for Causality	53
4.4	Results	54
4.5	Conclusion	59
4.5.1	Limitations	59
4.5.2	Future work	60

5	From Causality to Sarcasm: Mining the Shift of Opinions Regarding the Supreme Court on Twitter	61
5.1	Abstract	61
5.2	Introduction	62
5.3	Background	64
5.4	Methods	65
5.4.1	Overview	65
5.4.2	Data collection	66
5.4.3	Data cleaning	67
5.4.4	Analyses	69
5.5	Results	73
5.6	Discussion	76
5.7	Conclusion	77
6	Discussion, Future Work & Conclusion	78
6.1	Overview	78
6.2	Scope For Improvement & Current Exceptions	78
6.3	Conclusion	79
6.4	Availability of data and material	80
A	List of Abbreviations	81
B	Values For LDA Multicore model Parameters	82
	Bibliography	83

List of Tables

Table 2.1	Libraries used in each step (Section 2.4) of our experiments.	20
Table 3.1	Keywords used for topic ‘Obesity’ for Figure 3.5	38
Table 3.2	Versions of libraries used	40
Table B.1	LDA model parameters	82

List of Figures

Figure 1.1	Example of Obesity causal Map	2
Figure 1.2	Example of Politics causal Map	3
Figure 2.1	The Public Health Services Authority’s series of maps . . .	10
Figure 2.2	Our process in seven steps to validate a conceptual model using textual data.	15
Figure 2.3	Alternative view of our process, including libraries and APIs.	17
Figure 2.4	Typical pre-processing techniques applied to tweets. . . .	19
Figure 2.5	Pre-processing techniques applied to our Twitter data-set	19
Figure 2.6	Average number of edges confirmed using expert reports .	21
Figure 2.7	Average number of edges confirmed using Twitter data .	22
Figure 3.1	Proposed pre-processing techniques	30
Figure 3.2	Methodology explained step by step, including correspond- ing libraries and APIs used	31
Figure 3.3	CSV format for concept map to be fed to the pipeline . .	32
Figure 3.4	Causal Map - Obesity	36
Figure 3.5	Concept Map - Obesity	37
Figure 3.6	Comparison between synonyms and derivationally related form keywords	39
Figure 3.7	Computation results on CPU (Intel CORE i7 8th Gen) .	44
(a)	Time taken	44
(b)	Graphical representation	44
Figure 3.8	Computation results on GPU (nvidia 410.78)	45
(a)	Time taken	45
(b)	Graphical representation	45
Figure 3.9	Number of edges confirmed (out of 17 in Figure 3.5) for each combination of parameter values using Twitter data	46

Figure 4.1	Conceptual map	51
Figure 4.2	Data collection for Granger causality	52
(a)	Literature	52
(b)	Twitter conversation	52
Figure 4.3	Parameter tests for each hypothesis	55
Figure 4.4	Optimal Lag length selection	55
Figure 4.5	Causal Map using Granger causality	56
Figure 4.6	Causality plot of edge (weight discrimination, depression)	57
Figure 4.7	Causality plot of edge (obesity, stress)	58
Figure 4.8	Causality plot of edge (obesity, education)	58
Figure 4.9	Causality plot of edge (exercise, depression)	59
Figure 4.10	Causality plot of edge (intervention, exercise)	60
Figure 5.1	Conceptual model for our study	67
Figure 5.2	High-level view of our analysis process	70
Figure 5.3	Optimized values of our three parameters	72
Figure 5.4	Associations that were confirmed (green) or not (red) for both of the data-set	73
Figure 5.5	Sentiment analysis at both the node- and edge-level . . .	75
Figure 5.6	Prevalence of sarcasm at the association-level and at the sentiment-level	75

ACKNOWLEDGEMENTS

I thank the MITACS Globalink research award program for providing me with a great opportunity for learning and professional development. I am grateful for having a chance to meet so many wonderful people and professionals who led me through this research period.

I express my deepest gratitude and special thanks to my home supervisor Dr. Vijay Mago at Lakehead University, Canada and host supervisor Dr. Philippe J. Giabbanelli at Furman University, USA who in spite of being extraordinarily busy with their duties, took time out to hear, guide and keep me on the correct path and allowing me to carry out my project at their esteemed organization and extending during the training. Their guidance, constructive suggestions, and encouragement is the reason I was able to learn and grow as a researcher

I express my deepest thanks to my parents for taking part in giving necessary advice and guidance and arranging all facilities to make learning easier. I choose this moment to acknowledge their contribution gratefully.

I greatly appreciate the time taken by the committee members to review my thesis. I would also like to thank Datalab.science Team for helping with proofreading and revisions. MITACS, for their financial support provided through Application Ref. IT11172 and Funding Request Ref. FR27479. Publication costs were supported by an NSERC Discovery Grant for Dr. Vijay Mago. I thank Chetan Harichandra Mendhe for gathering the tweets under supervision of Dr. Vijay Mago. Without any of this, this research experience would not have been possible.

DEDICATION

I would like to dedicate this thesis to my parents, who gave me the foundation of something they always encouraged, education. For inspiring me and always believing in my ability to accomplish anything I set my mind to. For making sure I had access to all the resources I need to succeed no matter the circumstances. My younger brother, who always took care of my chores when I was busy studying and made me laugh whenever I would feel low. They are my driving force, what keeps me going.

I would also like to dedicate this thesis to all the teachers and mentors who inspired, empowered, taught, and shaped me into the person I am today.

Thesis Structure

This section provides an overview of the thesis research to show the reader where and how it validates the claims made in the abstract.

Chapter 1 introduces and explains the background, motivation and contribution of our research.

Chapter 2 discusses the impact of source selection on automatically validating complex conceptual models by comparing conceptual models derived from each source. Here the sources discussed are social media (Twitter in our case) and expert reports. The conceptual model used as a proof of concept is the one of Obesity.

Chapter 3 provides a comprehensive explanation of the methods used and their implementation by mining Twitter data to validate generic conceptual models.

Chapter 4 explains the approaches used to extend these conceptual models using Granger Causality by mining literature or Twitter.

Chapter 5 uses the conceptual model of Politics regarding the Supreme Court and Bret Kavanaugh event as a case study to discuss the shift of opinions on Twitter. In this chapter, first we use the proposed pipeline to validate the political conceptual model. Second we compare the sentiment analysis and sarcasm detection results to analyze if sarcasm can alter results of computational methods used to validate such models.

Chapter 6 discusses the overall conclusion of this research as well as the scope for improvement and current exceptions.

Chapter 1

Introduction

1.1	Background	1
1.2	Objectives	4
1.3	Main Contributions	5

1.1 Background

Researchers along with members of the public are actively involved in projects designed to learn from public participation[1] and have successfully addressed complex issues in science and society. Some have accumulated data with the help of thousands of bird watchers across North America to reveal trends in bird distributions and behaviors[2], whereas others have collected and prepared monitoring information to respond to pollution in order to address environmental degradation[3]. The public participation in scientific research (PPSR)[4] movement is much more than just the gathering of data for science or resource management by explicitly engaging the public in the research process. PPSR provides integrated outcomes from science, the individual participants, and social ecological systems, which makes such research more powerful in producing science-based knowledge with diverse understanding. Research on PPSR is also being conducted in fields like public health[5] (e.g., Cashman et al. 2008), where community-academic collaboration added value to the analytic and interpretive phases of research. This participatory environmental modelling[6] practice, in which scientists and mem-

bers of the public work together to develop conceptual models, is helping us reshape our definitions and understanding of broader socio-ecological systems as well as address environmental issues and provide solutions.

Participatory modelling results in structured models, often in the form of maps ranging from mind maps to concept maps and then causal maps. These maps are created using different software mapping tools[7] which are available under different names: concept mapping, mind mapping and argument mapping. These tools map and display complex information visually to aid and enhance research and learning. The names may be used interchangeably, but the choice of mapping tool largely depends on the purpose for which the tool is used. For instance, mind mapping allows us to imagine and explore associations between concepts. A causal map[8] is a particular type of concept or cognitive map. While a concept map indicates only that ideas or concepts are related in some way, a causal map illustrates the cause and effect relationships among concepts. On the map, these concepts are represented as nodes containing future issues, factors, events or outcomes, whereas the causal relationships between them are demonstrated in the graph by arrows [See Figure 1.1]. Existing models[9] have come up with the use of expert skeleton concept maps which are being used as a facilitative tool by schools and corporations[10] to emphasize meaningful learning and to efficiently assess ill-structured, problem-solving areas. These tools prove useful since concept maps are the basis of causal maps, and starting a concept map for a new knowledge domain [See Figure 1.2] can be a bit misleading without expert knowledge.

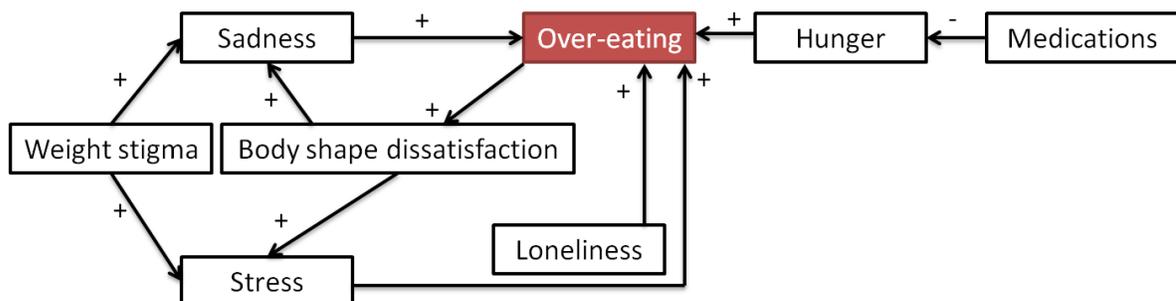


Figure 1.1: Example of Obesity causal Map

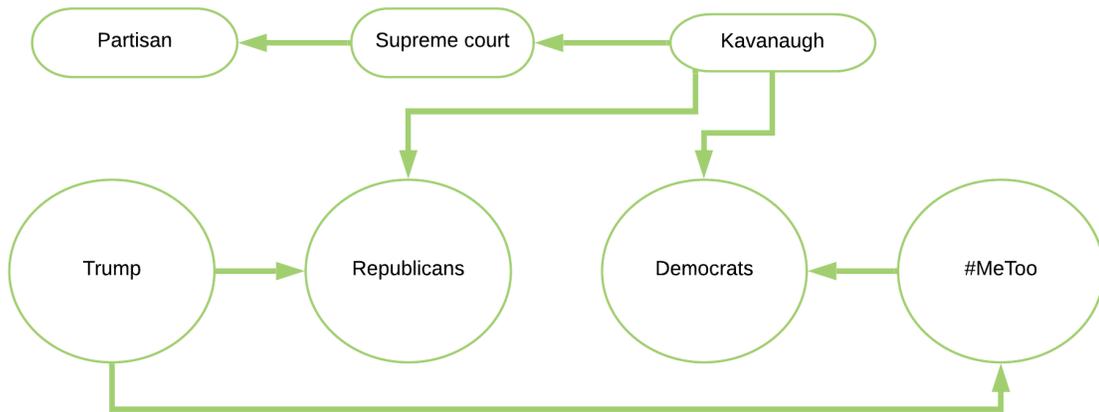


Figure 1.2: Example of Politics causal Map

There are limitations in participatory modelling: the power imbalance in eliciting perspectives, the logistics of bringing participants together, the problems of selection and representativeness, etc. It may not be too unrealistic to argue that the quality and the form of participation will be a driving factor in the success of future models, and that participation would be well served by the inclusion of cognitive, material and technological information[4]. In this research, we seek to address these limitations by developing causal maps that take into account the perspective of a very large number of participants. This will enable a move toward a citizen science approach[11] to causal mapping. The popularity of both citizen science and participatory modelling[12] has given way to a growing number of case studies, all of which outline the benefits of more inclusive forms of conservation planning. This public-science collaboration of participatory modelling and citizen science is often said to lead to the development of community-supported research and possibly improved environmental decision-making as well. Furthermore, the development of online modelling tools holds strong promise for the field of conservation biology.

Our approach proposes to take small causal maps (from participatory modelling processes) as a starting point and either confirm them with or extend them to the perspective of online participants. Specifically, we will draw on Twitter data, which has become an increasingly common approach to understanding perspectives on a large scale. Due to this remarkable increase in the use of social media, particularly from Twitter, researchers are able to obtain information about specific demographic communities[13] which are

difficult to reach through conventional means. Similarly, evidence suggests that Twitter can be used to raise public health concerns[14] and to anticipate complex contagions before they reach critical mass[15], thus pushing the platform towards being a reliable source for citizen science research.

Our overarching objective is to validate conceptual models with a citizen science approach, starting with ‘nucleus’ maps and using very large amounts of Twitter data to automatically confirm or extend these maps. This objective will be achieved by identifying a set of tweets relevant to a nucleus map and mining its themes. To demonstrate the feasibility of our approach, we will have published our work with different data-sets as a proof of concept in the next chapters because they represent the most frequently discussed public health and political issues on social media. By achieving this objective, we would be able to confirm the existing connections between two nodes. However, this approach is limited to establishing the associativity of concepts, and thus, it only produces undirected maps. Identifying causal connections, thus enabling directed maps, is an extension of the proposed work which we have implemented using Granger causality.

1.2 Objectives

Given the importance of developing accurate models of complex systems, the modelling process often seeks to be comprehensive by including experts and community members [16, 17, 18, 19, 20, 21, 22]. While many qualitative modelling processes can produce models in the form of maps [23] (e.g., cognitive/concept mapping, causal loop diagrams), they are generally conducted with a facilitator. Some of the limitations (e.g., costs, trained facilitator) may be addressed through emerging technologies [24]. However, one limitation remains: participants may not openly express their beliefs (e.g., weight discrimination or political opinions) when perceiving that they may not be well received by a facilitator or the research team. In contrast, the naturally occurring exchange of perspectives in social media provides an unobtrusive approach to collecting beliefs on causes and consequences of such complex systems. Mining social media may thus provide the views of community members [25, 26, 27, 28].

While obtaining a model via social media can inform policymakers about popular support for possible policies [29], the model may stand in stark contrast with an expert-based model [22]. Identifying and reconciling these differences is an important step to integrate

social computing (and specifically social web mining) with policy making. In this thesis, we contrast how mining social media instead of expert reports affects the validation of a large conceptual model of obesity. This overarching goal is achieved through three consecutive steps. First, we assemble a social media data-set (consisting of several million tweets) and several expert reports (totaling hundred of pages). Second, we employ an innovative multi-step process to examine a conceptual model using both the social media data-set and the expert reports. Third, we contrast the structure of these models using a Python package called NetworkX. This package creates, manipulates, and studies the structure, dynamics, and functions of complex networks. Subsequently, using sentiment analysis and sarcasm detection, we further confirm the connection based on computationally identified and categorized opinions expressed in tweets to determine whether the writer's attitude toward a particular topic, product, etc. is positive, negative or neutral. Finally using Granger causality on expert reports, we extend the concept map in terms of reverse causality.

1.3 Main Contributions

In the case of obesity, we found that three expert reports discussed 77% of all possibilities while millions of tweets on obesity and its cognates covered fewer interrelationships about 56.5%. Our methodology is generic as we provided proof of concept on four different data-sets and three different concept maps. The concepts analysed were Obesity and Politics since they are the most talked about on Twitter. The Twitter data used was different for each concept map. This proved that creating models using social media only may thus result in an oversimplification of complex problems. We explained our pipeline in detail with its optimized implementation and parameters used. Our proposed pipeline using GPU gave us two times faster results than our computations on CPU. Later we extended these conceptual models using Granger causality by mining literature. Our main focus was to mine Twitter conversations as an alternative which could be a leap forward. Furthermore, we used our pipeline to confirm conceptual models along with sentiment analysis and sarcasm detection to further see the nature of validation, that is whether people are talking positive, negative or neutral about a concept that we validated using our Twitter data. By automatically analyzing millions of tweets, we demonstrated that sarcasm in political tweets can significantly alter the outcome of tweet mining even when using large data-sets.

Chapter 2

From social media to expert reports: the impact of source selection on automatically validating complex conceptual models of obesity

2.1	Abstract	7
2.2	Introduction	7
2.3	Background	11
	2.3.1 Social web mining for health	11
	2.3.2 Conceptual models in obesity research	13
2.4	Validating a conceptual model from text	14
2.5	Comparing conceptual models from Twitter and expert reports	17
	2.5.1 Data-sets and pre-processing	17
	2.5.2 Validating the model for each data-set	19
2.6	Discussion	22
2.7	Conclusion	24

2.1 Abstract

Models are predominantly developed using either quantitative data (e.g., for structured equation models) or qualitative data obtained through questionnaires designed by researchers (e.g., for fuzzy cognitive maps). The wide availability of social media data and advances in natural language processing raise the possibility of developing models from qualitative data naturally produced by users. This is of particular interest for public health surveillance and policymaking, as social media provide the opinions of constituents. In this chapter, we contrast a model produced by social media with one produced via expert reports. We use the same process to derive a model in each case, thus focusing our analysis on the impact of source selection. We found that three expert reports were sufficient to touch on more aspects of a complex problem (measured by the number of relationships) than several million tweets. Consequently, developing a model exclusively from social media may lead to oversimplifying a problem. This may be avoided by complementing social media with expert reports. Alternatively, future research should explore whether a much larger volume of tweets would be needed, which also calls for improvements in scalable methods to transform qualitative data into models.

2.2 Introduction

Overweight and obesity are now a global phenomenon, found in economically developed or developing countries (e.g., United States [30], European countries [31], South Africa [32], China [33]) as well as in regions that experience a double burden with the concomitant problem of malnutrition [34]. While there are ongoing debates on a possible plateau or even decrease of overweight and obesity in the next generation, updated prevalence data for children suggests that severe obesity is on the rise [35]. There is a plethora of interventions to prevent overweight and obesity in both children [36] and adults [37], and an equally impressive number of interventions for treatment [38, 39]. Yet, individual struggles to achieve a health weight over a sustained period of time. For example, a review of weight management interventions found a weight loss over two years of 1.54 kg [40], which is far from the 5% weight loss recommended to produce health benefits [41]. These challenges have led to the realization that a simple solution would not suffice [42]: the health system needs to cope with the *complexity* of obesity [43, 44, 45].

The notion of complexity covers multiple characteristics, such as the vast individual differences (or *heterogeneity*) between weight-related factors [46, 47], or the nonlinear ways in which factors interact to form a system. The obesity system has been the subject of numerous studies [16, 48, 49, 50]. This system involves factors from a broad array of sectors (e.g., built environment, eating disorders, weight stigma [17, 51]), with interactions within as well as across sectors. Accurately modelling this system facilitates the development of integrated policies building on cross-sectoral efforts [52, 53]. If policies are developed separately along traditional themes (e.g., public planning works on the environment, doctors work on diseases and physiology, mental health experts work on psychology), then we have a heavily fragmented approach to obesity (Figure 2.1a). Efforts such as the Foresight Obesity Map [48, 54], or the Public Health Services Authority’s series of maps [17, 55, 56] thus support the development of synergistic policies working on integrated thematic clusters (Figure 2.1b).

Given the importance of developing accurate models of the obesity system, the modelling process often seeks to be comprehensive by including experts and community members [16, 17, 18, 19, 20, 21, 22]. While many qualitative modelling processes can produce models in the form of maps [23] (e.g., cognitive/concept mapping, causal loop diagrams), they are generally conducted with a facilitator. Some of the limitations (e.g., costs, trained facilitator) may be addressed through emerging technologies [24]. However, one limitation remains: participants may not openly express their beliefs (e.g., weight discrimination) when perceiving that they may not be well received by a facilitator or the research team. In contrast, the naturally occurring exchange of perspectives in social media provides an unobtrusive approach to collecting beliefs on causes and consequences of obesity. Mining social media may thus provide the views of community members [25, 26, 27, 28].

While obtaining a model via social media can inform policymakers about popular support for possible policies [29], the model may stand in stark contrast with an expert-based model [22]. Identifying and reconciling these differences is an important step to integrate social computing (and specifically social web mining) with policy making. In this chapter, we contrast how mining social media instead of expert reports affects the validation of a large conceptual model of obesity. This overarching goal is achieved through three consecutive steps. First, we assemble a social media data-set (consisting of several million tweets) and several expert reports (totaling hundred of pages). Second, we

employ an innovative multi-step process to examine a conceptual model using both the social media data-set and the expert reports. Finally, we contrast the structure of these models using network methods.

The remainder of this chapter is organized as follows. In section 2.3, we provide background information on the application of social web mining to health, and on the use of conceptual models in obesity research. In section 2.4, we briefly explain our approach to validate a conceptual model from text. In section 2.5, we perform this inference on both expert reports and tweets, and we examine how the conceptual models differ. Finally, these differences are discussed and contextualized in section 2.6.

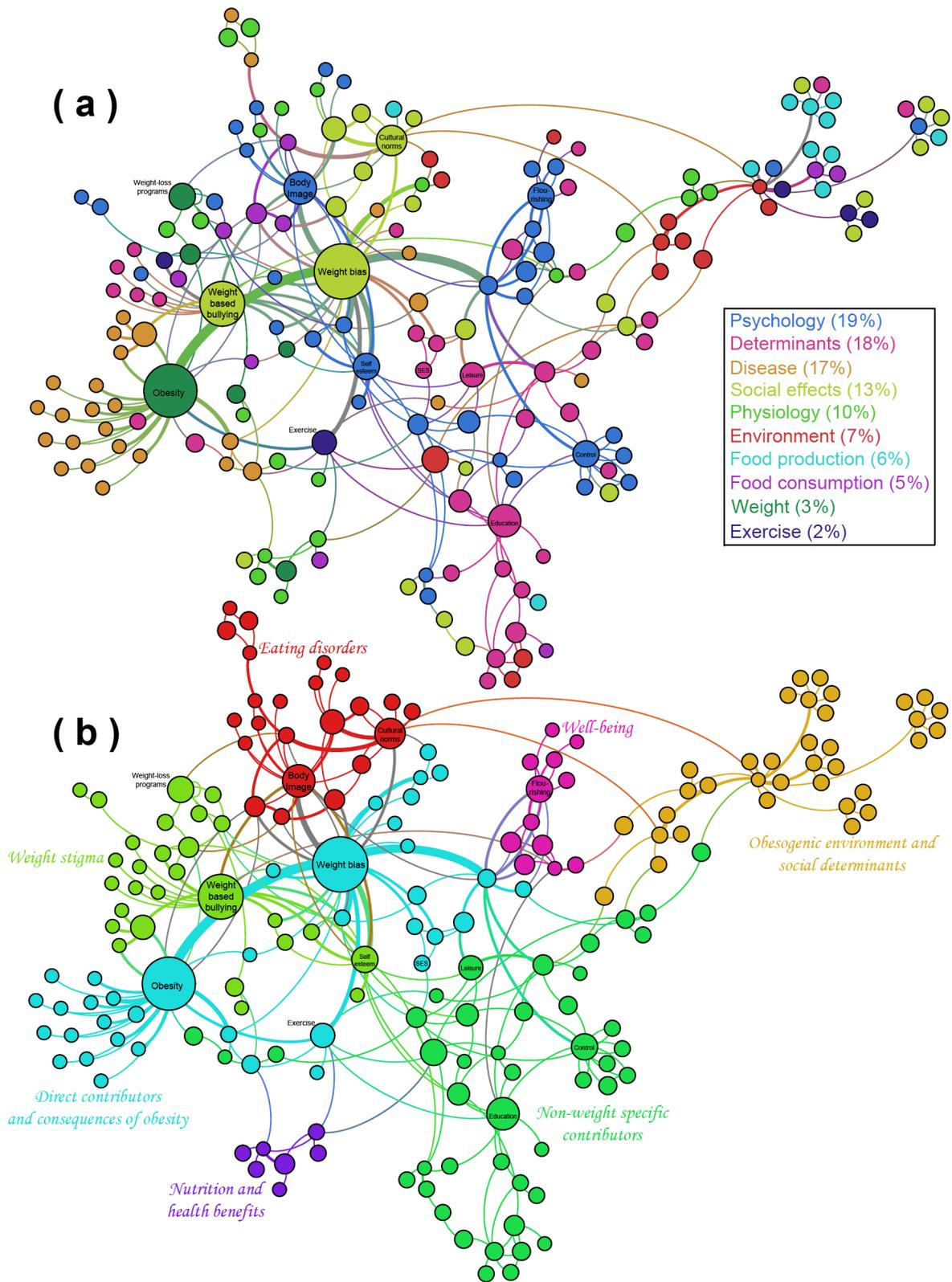


Figure 2.1: The Public Health Services Authority's series of maps [17, 55, 56] suggests that typical categories lead to fragmented approaches (a) whereas themes specific to overweight and obesity can support more integrated options (b). These maps are *conceptual maps* as they articulate how concepts (labeled circles) are related (curves).

2.3 Background

2.3.1 Social web mining for health

The social media of interest in this chapter is Twitter, in which users post and interact through short messages known as ‘tweets’. Twitter has been used for many studies on obesity and weight-related behaviors. For instance, Harris and colleagues collected 1,110 tweets and read them to understand how childhood obesity was discussed [57], while Lydecker *et al.* [58] read 529 tweets to identify the main themes related to fatness. Similarly, So *et al.* [25] analyzed the common features of 120 tweets that were most frequently shared (i.e., retweets) to understand what information individuals preferred to relay when it came to obesity. Reading the tweets to identify themes (i.e., content analysis) is a typical task to understand the arguments that a specific population uses on a subject of interest. Broader examples in health include the content analysis of 700 tweets [59] and 625 tweets [60] to examine the type of claims that health professionals make online, or an examination of 8,934 tweets documenting cyber incivility among nurses and nursing students [61]. While such content analyses make a valuable contribution to the body of knowledge on arguments in public health¹, they do not employ computational methods to automate (parts of) the analysis and thus scale it to a larger data-set. Automation can be as simple as counting how many times keywords of interest appear across tweets. Turner-McGrievy and Beets used `Hashtagify.me` to automatically count keywords in tens of thousands of tweets on weight loss, health, diet, and fitness. By dividing the analysis across time periods, they were able to examine if there are times of the year when individuals would be likely to consider weight loss, thus contributing to the timing of interventions [63]. Similarly, Sui *et al.* used the intensity of topics on Twitter as part of an effort to identify the public interest in intensive obesity treatment [64]. Such studies illustrate the important shift from having humans read and code all tweets to relying on a machine to handle most of a (much larger) data-set. The latter is the focus of data mining applied to the ‘social web’ (i.e. social web mining) which includes social networking sites such as Twitter but also encompasses blogs and micro-blogging. As Twitter has been the social platform of interest for many studies, the term of ‘Twitter mining’ has also emerged to refer specifically to the application of social web mining to Twitter [65].

¹While our focus is on analyzing the *text* provided by tweets, studies on Twitter that are primarily human- rather than computer-based are not exclusively content analyses. In the study of May *et al.* [62], the researchers created Twitter accounts for fictional obese and non-obese characters. They evaluated whether the weight status mediated how other users would interact with them.

Social web mining started to garner attention in the late 2000's to early 2010's. The application of social web mining to health was discussed in 2010 by Kamel Boulos *et al.* [66] and in 2011 by Paul and Dredze [67], showing how a broad range of public health applications could benefit from mining Twitter. Studies have been able to mine a staggering volume of data, going well over what a team of humans could handle. For example, Eichstaedt *et al.* mapped 148 million tweets to counties in an effort to relate language patterns to county-level heart disease mortality [68]. At an even larger scale, Ediger and colleagues used a Cray computer to approximate centrality within two hours on a data-set of interactions between Twitter users comprising 1.47 billion edges [69]. While these cases are noteworthy by their volume of data, studies employing social web mining for obesity research typically involve millions of tweets². Using 2.2 million tweets, Chou and colleagues found that tweets (as well as Facebook posts) often stigmatized individuals living with overweight and obesity [26]. In two studies on obesity and weight-related factors, Karami analyzed 6 million [27] and 4.5 million tweets [28]. In a study of health-related statistics, Culotta mined 4.3 million tweets and found that the data was correlated with obesity [71]. Given that obesity is driven by many factors (e.g., eating behaviors, physical activity behaviors), there is also a wealth of large-scale studies on such factors, such as the work of Abbar *et al.* on 503 million tweets regarding food [72]. Finally, the value proposition of several new platforms is not the analysis of one particular data-set, but rather the ongoing ability to monitor diet or physical activity. This is particularly the case for the Lexicocalorimeter, which measures calories in each US state via Twitter [73], and to a lesser extent for the National Neighborhood data-set of Zhang *et al.* which tracks diet and physical activity through Twitter [74].

Several commentaries [75] and reviews [76, 77, 78] have explored whether this abundance of studies has contributed to public health. Findings depend on what specific aspect of health is concerned. Social media has yet to impact practices in public health surveillance [77], but a review centered on chronic disease found a benefit on clinical outcomes in almost half of the studies [76], and a review specific to obesity highlighted a modest impact on weight [78].

²There are several exceptions of studies employing smaller data-set. However, their objectives may not be to identify themes (which necessitates a large volume of tweets), thus they can accomplish their goals with a smaller data-set. A case in point is the work of Tiggemann and colleagues, who used 3,289 tweets to examine interactions between Twitter communities that promoted either a 'thin ideal' or health and fitness [70].

2.3.2 Conceptual models in obesity research

Although our work will involve the identification of themes, we have a very different endeavor from studies reviewed in the previous section, which focused on identifying themes and their variations across time, places, or communities of users. Our objective is to contrast *conceptual models* that have been automatically extracted from tweets and expert reports. As evoked in the introduction, models of complex systems such as obesity support several important policy-making and analytical tasks. In this section, we briefly review the features that models often seek to capture when it comes to complex health systems, and how models are used in obesity research specifically. Penn detailed key characteristics of complex health systems that justify the development of models (emphases added):

“Many problems that society wishes to address in population health are clearly problems of *managing complex adaptive systems*. They involve making interventions in systems with *multiple interacting causal connections*, which *span domains* from physiological to economic. Additionally, of course, the individuals whose health we ultimately wish to improve adapt and *change their behavior* in response to medical or policy interventions.” [79]

Several of these points were echoed by Silverman in justifying the use of systems-based simulation for population health research [80]. modelling changes in the heterogeneous health behaviors of individuals often uses the simulation technique of Agent-Based modelling, and has been done in obesity research on multiple occasions [81, 82, 83, 84, 85]. Such models can be very detailed and use widely different architectures to capture the cognitive processes of the agents. Validating them using text is thus an arduous task. modelling interacting causes across domains has been achieved in obesity research through a variety of techniques. System Dynamics (SD) allows to represent nonlinear interactions between weigh-related factors over different time scales and at different strengths [86, 87]. However, much like agent-based modelling, the great level of details supported by SD makes it difficult to derive or validate such models from text. Fuzzy Cognitive Maps (FCM) are a simpler alternative that eliminates the notion of time to focus on the different strengths of causal relations [22, 88, 89, 90]. Such models can be compared [22], but validating them from text still requires a trained analyst [91]. An even greater simplification is to use conceptual rather than simulation models. Conceptual models cannot run scenarios or what-if questions, and cannot ‘generate’ numbers. Instead, their focus is to capture relevant factors and whether they are connected [92].

Conceptual models can be compared [93] and validated using text as shown in our previous work [92].

There are several types of conceptual models [23]. We recently detailed the differences between causal maps, mind maps, and concept maps [24]. In short, this chapter focuses on concept maps (Fig. 2.1), which are undirected networks representing concepts as nodes and relationships as edges. Similarly to the other forms of conceptual models aforementioned, a concept map supports policy-oriented tasks such as identifying clusters [54] (e.g., to coordinate actors across domains on one problem such as food) or finding feedback loops [17, 55, 56] (e.g., to use as leverage points in an intervention).

2.4 Validating a conceptual model from text

The process starts with a conceptual model that we seek to validate, and the text corpus is used to validate. Intuitively, our process uses the concepts' names to find relevant parts of the corpus and find which concepts tend to co-occur. Technical aspects include handling variations in language (as we cannot rigidly assume that a concept's name will appear as such), identifying themes, and mapping themes from the corpus back to concepts in the conceptual model. Our process uses seven steps, illustrated on a theoretical example in Figure 2.2. The first two steps are performed for each concept node:

- (1.a) We replace all concepts' names and words from the corpus with their base form (i.e., lemma). This is accomplished through *lemmatization*, which uses a morphological analysis to remove inflectional endings. This step ensures that minor variations of a term are all mapped to the same one (e.g., 'flooding' and 'floods' are all mapped to 'flood').
- (1.b) Each lemmatized concept names is expanded with derivationally related forms. For instance, instead of only searching for 'flood' in the corpus, we will also accept words such as 'deluge'.
- (2) For each concept (i.e., the expanded lemma), we retrieve all parts of the corpus that contain it. For instance, the concept 'flooding' will lead to retrieving all tweets include the lemmas 'flood' or 'deluge'.

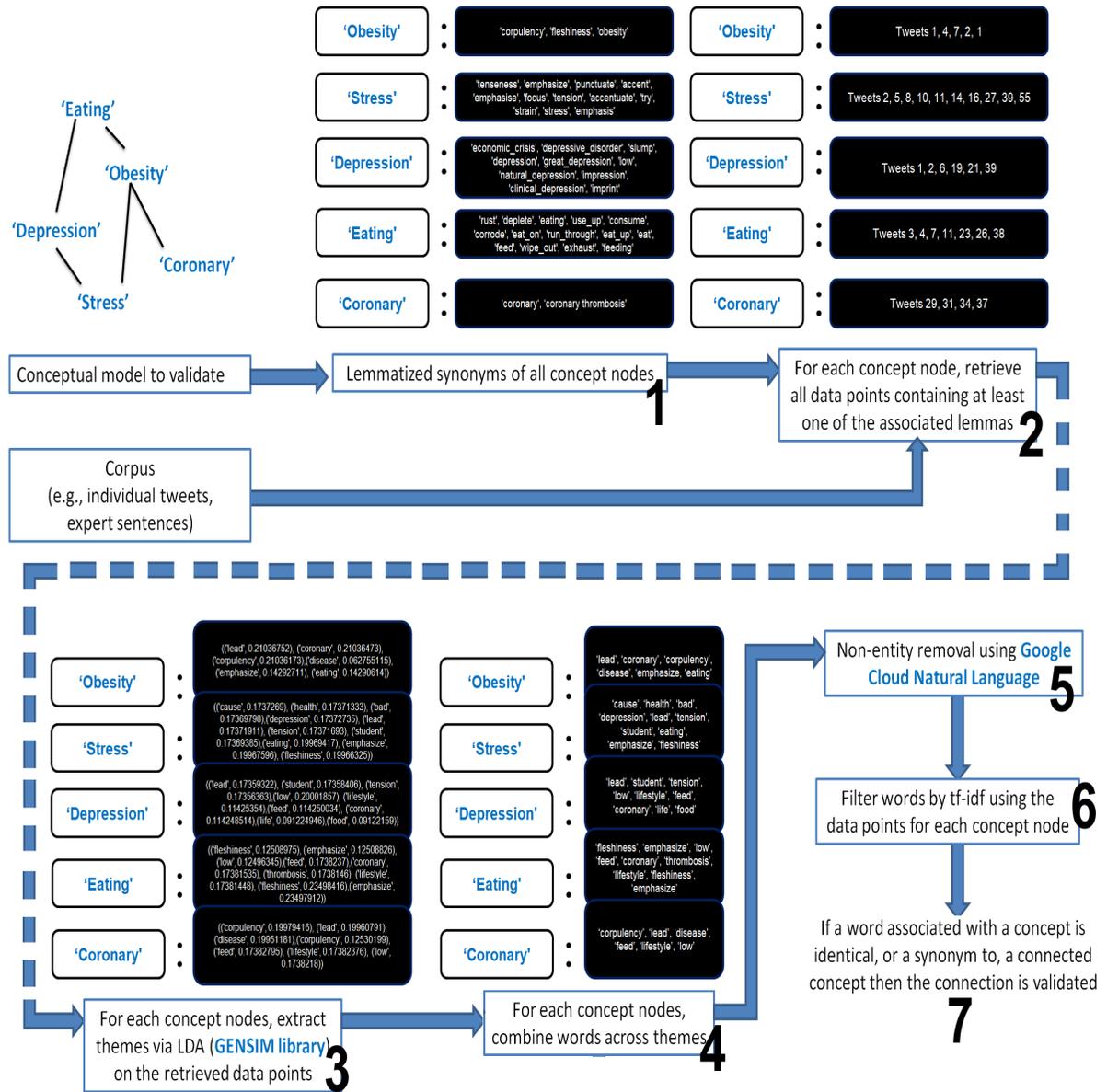


Figure 2.2: Our process in seven steps to validate a conceptual model using textual data. *The high-definition figure can be zoomed in for details.*

Upon completion of step 2, we have related a portion of the corpus to each concept node. We then find the themes in each portion of the corpus using three parameters:

- (3) We apply the Latent Dirichlet Accuracy (LDA) model to find prevalent themes. The two parameters for this step are the number of themes and number of words per theme.
- (4) We gather words across themes into a single set of words. This set is cleaned by

removing words that are already present in the set of derivationally related form of the node. In other words, we only look for concepts that the node could be *associated* with but not *equivalent* to.

- (5) Since concepts' names are entities, a concept can only be associated with an entity. Consequently, we remove all non-entities from the words.
- (6) At this step, we have a set of entities that a concept node could be associated with. However, some of the entities may be noise rather than meaningful associations. We thus sort the entities by tf-idf (term-frequency inverse-document-frequency) computed over the set of tweets in which each word appears. We use a threshold parameter to identify which entities have a sufficient tf-idf to be selected.

Upon completion of step 2, we found entities that a concept node could be associated with. The final step goes back to the conceptual model to see if the association exists:

- (7) For each node, we compare its associated entities with its connected nodes and derivationally related forms. If there is a match, then the text corpus has confirmed an association between the two concepts. If no match is found, the association is not confirmed. Note that associated entities that do not match any connected nodes suggest additional connections, which is a different from validation as we seek to confirm existing connections.

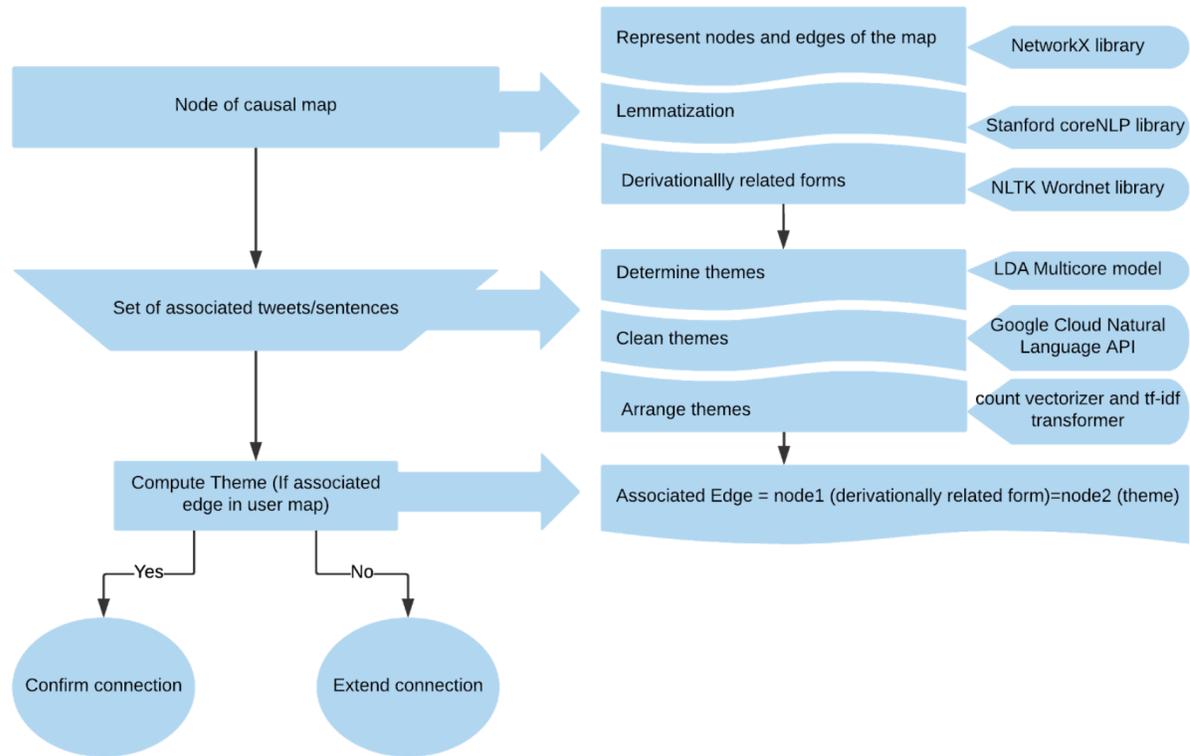


Figure 2.3: Alternative view of our process, including libraries and APIs.

This process is also depicted in Figure 2.3, listing the libraries that can be used for each step. The specific versions of the libraries used in our experiments are included in section 4.

2.5 Comparing conceptual models from Twitter and expert reports

2.5.1 Data-sets and pre-processing

The conceptual model that we seek to validate was developed with the Provincial Health Services Authority (PHSA) of British Columbia to explore the interrelationships involved in obesity and well-being. The model was presented in 2015 at the Canadian Obesity Summit [17] and tested with policy makers in 2016 [56]. The model is now part of the `ActionableSystems` tool [55] can be downloaded at <https://osf.io/7ztwu/> within ‘Sample maps’ (file *Drasic et al (edges).csv*). The model consists of 98 nodes and 177 edges. From here on, we will refer to it as ‘the PHSA map’.

To validate the PHSA map, we used two data-sets. Our first data-set (‘the Twitter data-set’) consists of 6,633,625 tweets in the English language on obesity collected from Oct. 2, 2018 to Oct. 4, 2018. The number of tweets was chosen to be in line with comparable studies at the interface of natural language processing and obesity research [26, 27, 28]. The keywords to collect the tweets included each of the 98 concept names in the PHSA map as well as their synonyms automatically retrieved through WordNet. For instance, we used not only ‘obesity’ but also words such as ‘fatness’, ‘corpulent’, ‘embonpoint’ and ‘fleshiness’. Similarly, physical activity was expanded to include many forms such as calisthenics, isometrics, jogging, jump rope, and so on. The rationale is that the map contains abstract concepts, but individuals may speak of specific instances or use a variety of words to describe the same abstraction.

After collecting a large number of tweets, natural language applications require extensive pre-processing. The impact of each options (and their interactions) on results obtained from Twitter has been extensively described when performing sentiment analysis [94, 95, 96] and in more generic tasks such as classification [97]. Some of these options are summarized in Figure 2.4 and include the removal of parts deemed unnecessary for analysis (e.g., hashtags, URLs, numbers, non English words) or the mapping of data into forms that can be more conveniently processed (e.g., expanding acronyms and abbreviations, replacing emojis, spell checking). The pre-processing options used for our data-set are depicted in Figure 2.5. These options are chosen specifically for our research question: for instance, we remove stop words because they cannot be meaningful concept names in a model, but other analyses (e.g., attributing tweets to specific writers) may have kept such words. The order of the steps also matters: for instance, we cannot perform part-of-speech tagging and lemmatization (step 5) before ensuring that all the words have been corrected (step 3). After pre-processing, our data-set included 1,791,333 tweets.

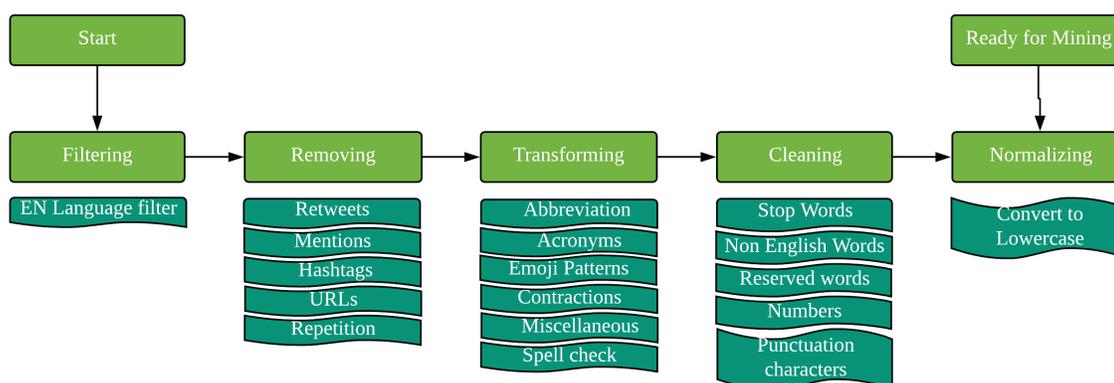


Figure 2.4: Typical pre-processing techniques applied to tweets.

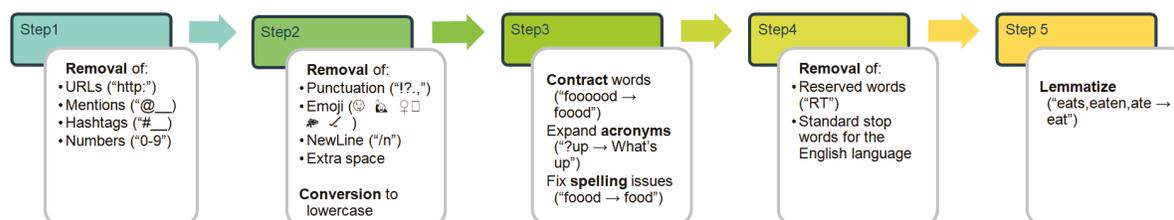


Figure 2.5: Pre-processing techniques applied to our Twitter data-set in a specific order. We used a Spell Checker library in step 3, the Natural Language Toolkit (NLTK) for steps 1-4, and the Stanford coreNLP library for step 5.

The second data-set is formed of three reports on obesity: the 2010 report from the white house task force on childhood obesity [98], the 2013 report to the Provincial Health Services Authority [99] and its 2015 update (whose findings are published in [17]). We combined the three reports with the PyPDF2 library, leading to 310 pages, and we kept 247 pages after removing those that were either blank or only contained images. Pages were then transformed into raw text using the pdftotext library and divided into 4,302 sentences using the full point ('.'). Pre-processing was finally applied, using the same script as for tweets while noting that several options such as removing emojis would not be triggered. The resulting data-set had 3447 sentences.

2.5.2 Validating the model for each data-set

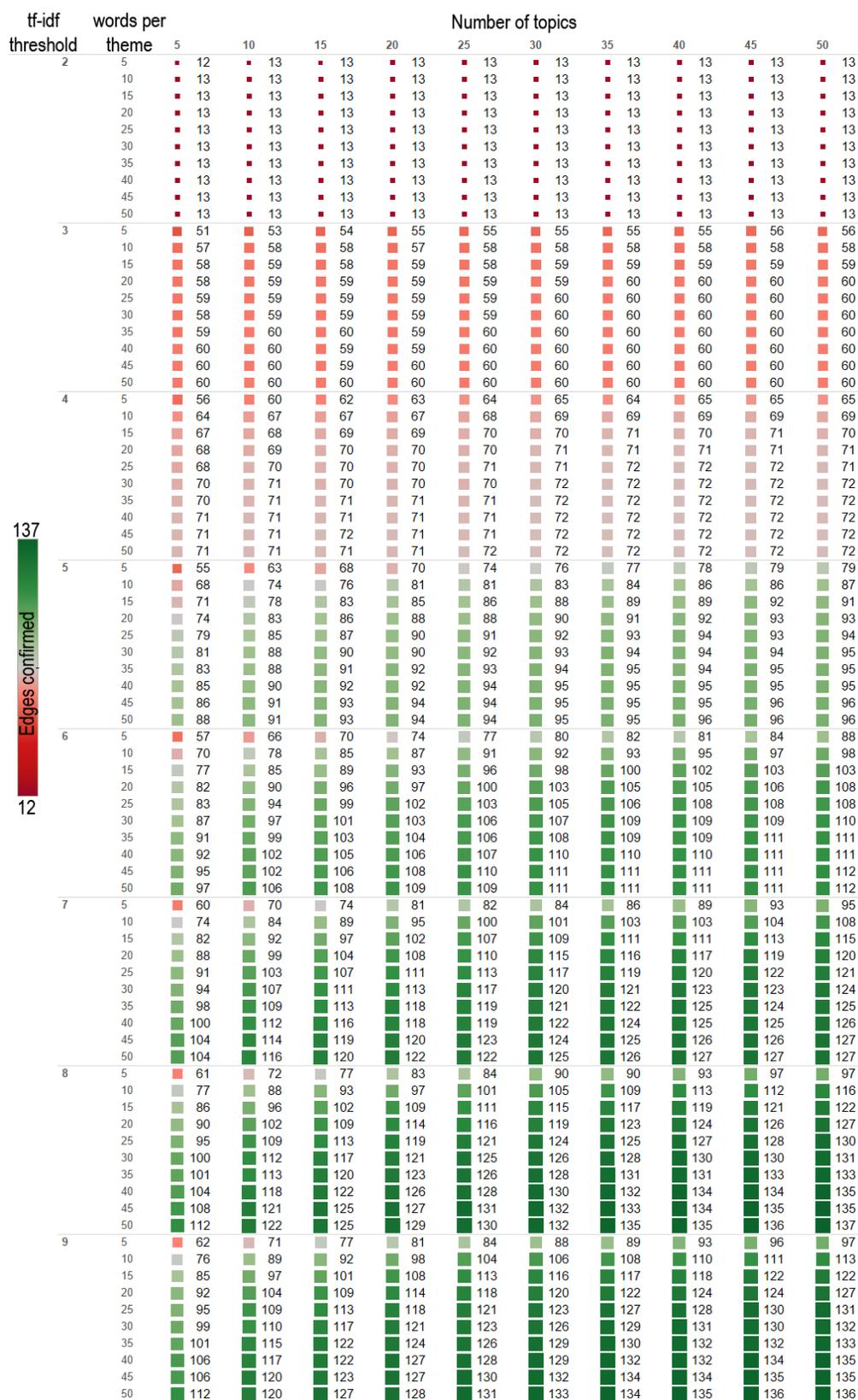
The methods introduced in section 3 are implemented in Python, relying on libraries as listed in Table 2.1. While our implementation was able to cope with millions of tweets,

we note that a larger volume of data may also require a distributed database architecture and an efficient search engine such as Elasticsearch [100].

Step	Library	Used for
	NetworkX	Conceptual model (accessing node labels and edges)
1a	Stanford coreNLP	Lemmatization
1b	WordNet	Derivationally related forms
3	[101]	Parallel, multi-core Latent Dirichlet Allocation (LDA) model for big data
5	Google Cloud Natural language API	Entity identification
6	scikit-learn (CountVectorizer, TfidfTransformer)	Sorting words by tf-idf

Table 2.1: Libraries used in each step (Section 2.4) of our experiments.

Our approach has three parameters: number of themes, number of words per theme, and tf-idf threshold to eliminate noise. Hyperparameter optimization was thus necessary to use each data-set most efficiently, and fairly compare their potential in validating a model. To optimize performances with expert reports, we performed a grid search by varying the number of topics and words per topic from 5 to 50 in increments of 5, and we varied the tf-idf from 2 to 9 by increments of 1. This resulted in 800 combinations of parameter values. As there is randomness in the LDA model, we performed ten experiments per combination of parameter values, leading to a total of 8,000 experiments. At most, our process validated an average of 136.5 edges (77.11% of the map) using 50 topics, 50 words per topic, and a td-idf threshold of 8 (Figure 2.6).



137
Edges confirmed
12

Figure 2.6: Average number of edges confirmed (out of 177 in the PHSA map) for each combination of parameter values over ten experiments using expert reports

A grid search was also performed on the Twitter data-set. However, our current implementation takes approximately five days to compute the results for one combination of parameter values (single experiment), using a server-grade workstation (Dual Xeon Gold 6140). Given this limitation, we used single experiments and a coarser grid. To optimize performances with Twitter data, we performed a grid search by varying the number of topics and words per topic from 5 to 50 in increments of 15, and we varied the tf-idf from 2 to 9 by increments of 2. This resulted in 64 combinations of parameter values. As there is randomness in the LDA model, we performed three experiments per combination of parameter values, leading to a total of 192 experiments. At most, our process validated 100 edges (56.5%) using 50 topics, 50 words per topic, and a tf-idf threshold of 9 (figure 2.7).

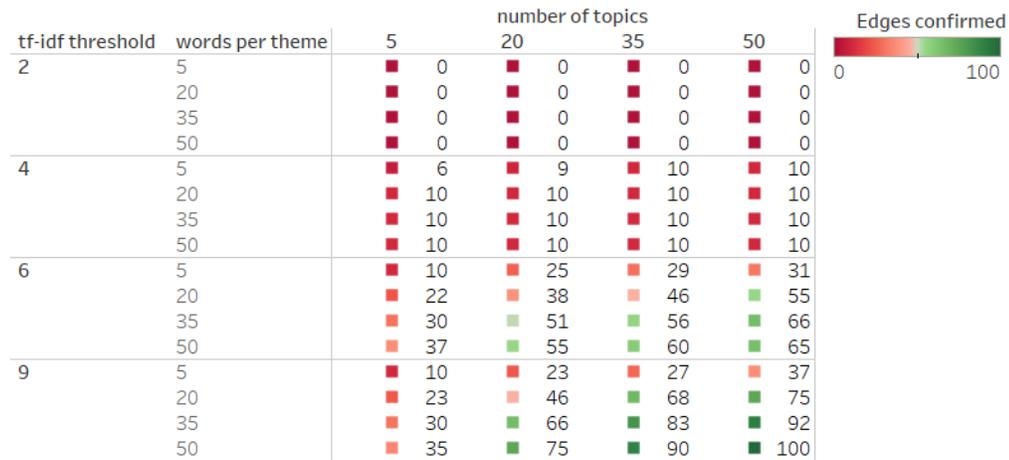


Figure 2.7: Average number of edges confirmed (out of 177 in the PHSA map) for each combination of parameter values over ten experiments using Twitter data

2.6 Discussion

A focus group with a few participants may only discuss some of the interrelationships at work in overweight and obesity, and may avoid sharing opinions that are potentially disapproved by others. In contrast, social media such as Twitter provide access to a massive number of participants who can use conditions of anonymity to share opinions more freely. Social web mining applied to Twitter thus comes with the potential to explore many interrelationships in an unobtrusive fashion. In particular, crowdsourcing over Twitter holds the promise of easily building large conceptual models, under the assumption that at least some groups of users will touch on each part of the model. Our

study questions this potential and promises by analyzing whether millions of tweets are more useful to develop a conceptual model of obesity than a handful of reports.

Although conceptual models can be automatically compared [93], developing a model from each data-set (tweets vs. reports) and comparing them would not be able to tell us which one is ‘better’. Our study question thus requires a referential. We use a previously developed conceptual model of obesity and well-being to serve as referential, and we establish how much of this model would have been obtained if we used either tweets or reports. In other words, we measured the percentage of the model’s structure that is confirmed with each data-set.

While both data-sets were able to cover over half of the model, we note that it only took three expert reports compared to using millions of tweets. In addition, despite the abundance of tweets, the three expert reports touched on more relationships. Within our application context, these results suggest that an exclusive reliance on social media may result in oversimplifying a complex system, thus limiting the potential to automatically develop models using such a source. We note that a comprehensive analysis across subjects and using a variety of maps would be needed to assess whether our results produced on one model (the Provincial Health Services Authority map) and one application subject (obesity) can be generalized to other models and subjects.

There are several limitations to this study, which we intend to address in our future research. First, one of the premises of big data research is that a large volume may compensate for many imperfections in the individual data points. Although we used a similar number of tweets to other studies at the interface of natural language processing and obesity research [26, 27, 28], it is possible that some of the interrelationships of the model we seek to validate are rare and thus only detectable in even larger data-sets. Repeating this study with significantly larger data-sets could elucidate this question. However, we then run into the second issue: our process to validate a causal map against textual data is very computational intensive. The search space to optimize the result is defined by three parameters which involve randomness, thus requiring several experiments for each combination of parameter values. On a server-grade workstation, a single combination with a CPU-based implementation requires in the order of days. Optimizing results and using larger data-sets will thus require implementations that scale, with a particularly promising option consisting of a GPU-based implementation. Alternatively, we may re-

duce the search space if we can better characterize the impact that parameters generally have on the results and then devise more computational efficient processes. For instance, the tf-idf threshold plays an essential role in driving performances (Figure 2.6) but may be replaced by additional pre-processing steps preventing the inclusion of noise, such as classifiers removing unwanted documents [102].

2.7 Conclusion

Both social media data and expert reports may be used to take into account popular perspectives and expert opinions when creating large conceptual models. In the case of obesity, we found that three expert reports discussed 77% of all possibilities while millions of tweets on obesity and its cognates covered fewer interrelationships about 56.5%. Creating models using social media only may thus result in an oversimplification of complex problems. In our next chapter we discuss our pipeline in detail with its optimized implementation and parameters used. ¹

Notes

¹This chapter is published in the proceedings of the 11th International Conference on Social Computing and Social Media [103]

Chapter 3

Validating conceptual models by mining Twitter data

3.1	Abstract	26
3.2	Introduction	26
3.3	Methods	29
3.3.1	Overview	29
3.3.2	Representing nodes and edges of user concept map	31
3.3.3	Mapping derivationally related form onto concept nodes in lemma form	33
3.3.4	Mapping relevant tweets onto concept nodes	34
3.3.5	Extracting themes from mapped relevant tweets	35
3.3.6	Validating edges of user concept map	36
3.4	Experiment	36
3.4.1	Experimental set-up	36
3.5	Implementation	40
3.6	Discussion	43
3.7	Conclusion	45

3.1 Abstract

A model can represent anything in its scope either by using physical objects which are sometimes difficult to verbalize or by using concepts or ideas which are based on interpreters verbal and cognitive ability. Furthermore, a model made of ideas or concepts instead of physical objects is called a conceptual model. Conceptual models are used by practitioners and researchers to facilitate meaningful frameworks in decision-making activities. These models identify the key components and structure them into a systematic and consistent workflow. These models are created using in-depth qualitative interviews or questionnaires. Twitter is already emerging as a good source of data for content analysis, surveillance, engagement and network analysis in public health-related fields. Consequently, developing a conceptual model by complementing social media with expert reports may lead to oversimplifying a problem. Our previous work contrasts a model produced by social media with one produced via expert reports. It also provides a proof-of-concept that conceptual models could be validated using Twitter. The results are confirmed through day to day exchange of information or ideas on that particular subject matter on Twitter as well as through expert reports or research papers related to that particular subject model. While this paper proposes a pipeline used to automatically validate these conceptual models by mining Twitter data using machine learning approaches.

3.2 Introduction

A concept map shows relationships among concepts, while concept mapping is a method of visualizing these defined relationships between concepts. In the early 1970s, Novak [104] developed knowledge representation using concept mapping for science education and has since been implemented for research, education or evaluation. Concept maps in science are used to categorize and organize concepts, show hierarchy and present inter-relationships between the concepts. They have been used as assessment tools to capture the mental models [105] that learners use when building them. The use of conceptual models is expanding with systems becoming more complex [106]. With this expansion, the model's effectiveness to cover the system's fundamentals is being realized, leading to the development of numerous conceptual modelling techniques. Complementing concept mapping with conceptual modelling helps to understand complex phenomena [107].

Conceptual models are used in cognitive psychology and philosophy in the form of mental models which could be a representation of ideas or concepts in one's mind [108] or a model of the mind itself [109]. These models also exist as mathematical models which can take forms of statistical model, differential equations or game theoretic models. Correspondingly, conceptual models in economic sector [110] are represented by a set of variables and a set of logical and quantitative relationships between them to forecast economic activity, propose economic policy or present reasoned argument to justify these policies. Then comes the models in information system design [111] where conceptual models of human activity systems are used as a system analysis method [112] concerned with problem structuring in management. They are used in software engineering for requirement analysis [113] as well as for representation of data in entity-relationship model [114] form. In health education research and practices, the conceptual model developed guides health education research or practice as a diagram of proposed causal linkages among a set of concepts believed to be related to a specific public health problem [115]. Conceptual models help in narrowing both research questions and the targets of intervention. Highly complex domains like humanities and social sciences use appropriate conceptual modelling tools [116] to explore, understand, document and communicate such domains. Even tools like tablets are being used to support entity-oriented exploratory search in knowledge graphs [117]. Research on theories of concepts underlying conceptual modelling, methods, and tools for developing and communicating conceptual models, techniques for transforming conceptual models into effective implementations, and the impact of conceptual modelling techniques on databases, business strategies and information systems development, etc. [118] is effective. The research topics include conducting entity search over knowledge bases, exploiting conceptual modelling in data crowd-sourcing, enterprise data management, and data quality control, to name a few [119] [120]. Our approach focuses on the validation of a complex system using conceptual modelling by mining social media as a knowledge base for validation to oversimplify a system and its problem.

The validation of conceptual models are usually done by stakeholders [121], pre-defined criteria [122] or expert approach [123] which depends completely on human experience, knowledge, and clarity of visual representation of the map. Validation of a conceptual model is important as it validates that the theories and assumptions underlying the conceptual model are correct to aid in decision-making [124]. A conceptual model could be validated using a subjective approach by exploring model behaviour or objec-

tive approach by doing a comparison using statistical tests [124]. On the other hand validation of conceptual model via social media gives incite of the popular public supported opinions [29]. Considering that we selected Twitter as a social media platform where people exchange everyday views and information for validation of generic concept maps. This platform has been used for various studies related to health practices [58] [57]. Our previous work [103] contrasts a model produced by social media with one produced via expert reports. It also provides a proof-of-concept that conceptual models could be validated using Twitter. Validation here means that our proposed approach confirms the association between the concepts of a specific topic-related model based on the day-to-day communication for that specific topic under discussion on Twitter as well as expert reports or research articles. It concludes that both social media data and expert reports may be used to take into account popular perspectives and expert opinions when creating large conceptual models. For proof-of-concept we validated accurate model of obesity system which was presented in 2015 at the Canadian Obesity Summit [17] and tested with policymakers in 2016 [56], where we found that three expert reports discussed 77% of all possibilities while millions of tweets on obesity and its cognates covered fewer interrelationships. Creating models using social media only may thus result in an oversimplification of complex problems.

In this paper we propose a pipeline used to automatically validate these conceptual models by mining Twitter data using machine learning approaches. This paper gives a better understanding of what methods were used, what parameter values to set, how were they implemented and how did we improve our implementation.

3.3 Methods

3.3.1 Overview

Algorithm 1: Validating conceptual map

Input : user concept map in CSV format as *mapFile.csv*

Twitter data as D

Output: Validated concept map as G'

```

1  $G \leftarrow initialize(mapFile.csv)$ ; //  $G = (V, E)$  using NetworkX
2 for  $V_i \in G$  do
3    $der\_rel\_form[V_i] \leftarrow getDerivationallyRelatedform(V_i)$ ; // using WordNet
   (NLTK)
4    $relevantTweets[V_i] \leftarrow retrieveTweets(der\_rel\_form[V_i], D)$ ; // mapping
   vertices to the tweets that contain them using their
   derivationally related form
5    $relevantThemes[V_i] \leftarrow extractThemes(relevantTweets[V_i])$ ; // using
   gensim LDA multicore model, google cloud natural language API,
   scikit-learn count vectorizer and tf-idf transformer
6   for  $k$  in  $relevantThemes[V_i]$  do
7     for  $j$  in  $der\_rel\_form[V_i]$  do
8       if  $der\_rel\_form[V_{ij}] = relevantThemes[V'_k]$  then
9          $E' \leftarrow [(V_i, V'_i)]$ ; // mapping together vertices with similar
           derivationally related form and extracted themes
           keyword to generate an associated edge  $E'$ 
10        if  $E' = E$  then
11           $G' \leftarrow E'$ 
12        end
13      end
14    end
15  end
16 end
17 return  $G'$ ; // Finally the pipeline returns the validated concept map

```

The pipeline for automatically validating large conceptual models of complex systems using a large text data-set (social media such as Twitter) includes technical innovation in applying machine learning using NLTK's WordNet interface, Gensim's LDA multicore

model, Google Cloud Natural language API and, scikit-learn's count vectorizer and tf-idf transformer. In other words, the combination of all these machine learning approaches will help us confirm complex conceptual models with less hassle of money, time and resources for policy making.

Given two inputs (pre-processed Twitter data-set explained in section Experiment using process shown in Figure 3.1 and conceptual model shown in Figure 3.5), our pipeline generates the output (validation of the conceptual model). Here the pre-processed Twitter data-set as well as the conceptual model used are explained in subsection Experimental set-up. See Algorithm 1 for the overall algorithm we used to create our pipeline to validate conceptual models.

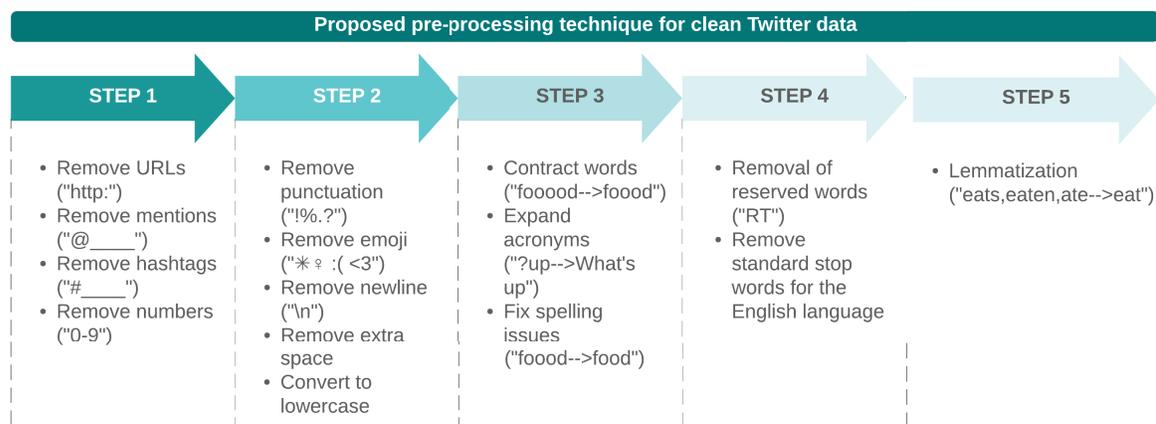


Figure 3.1: Pre-processing techniques applied to our Twitter data-set in a specific order. We used a Spell Checker library in step 3, the Natural Language Toolkit (NLTK) for steps 1-4, and the Stanford coreNLP library for step 5.

In the next subsections, we explain our methodology step by step according to the proposed approach (see Figure 3.2) using the experimental setup explained later.

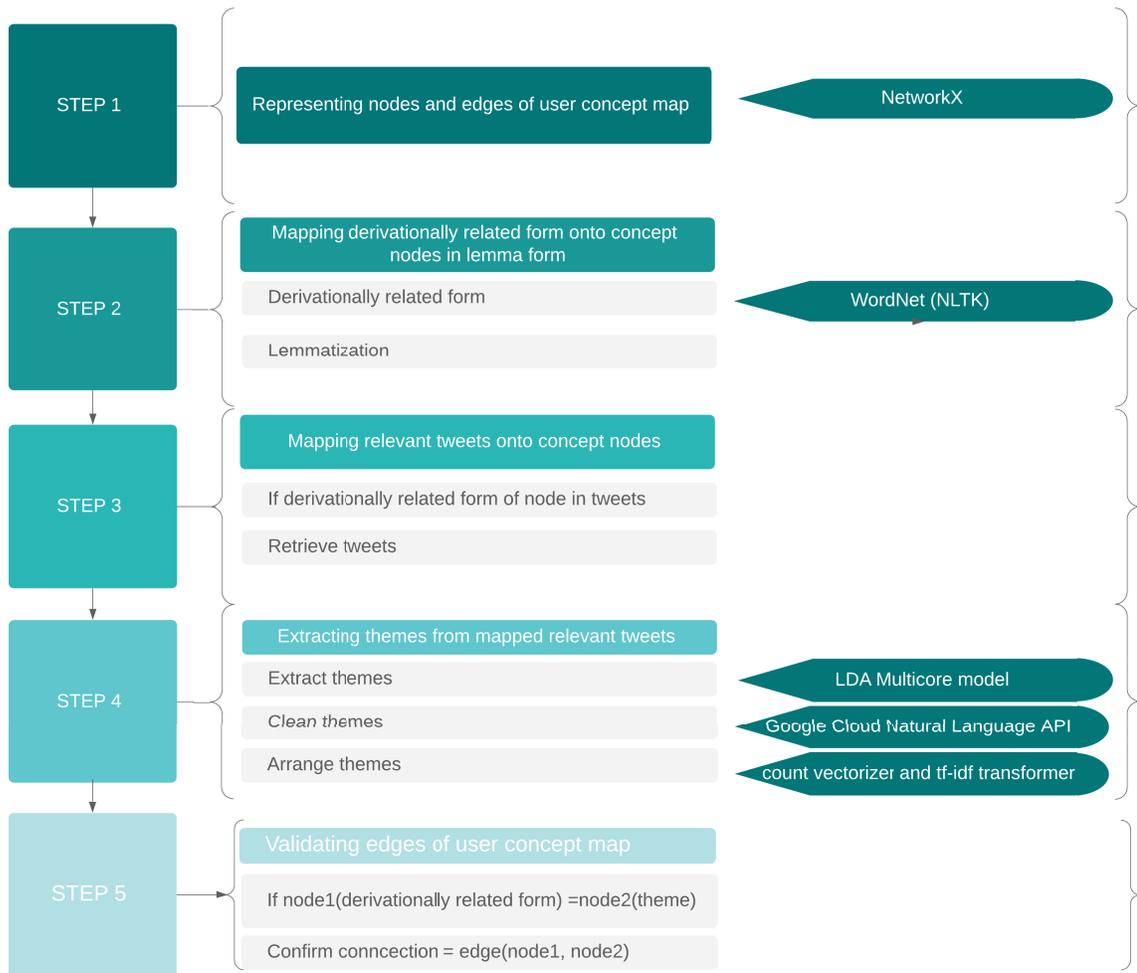


Figure 3.2: Methodology explained step by step, including corresponding libraries and APIs used

3.3.2 Representing nodes and edges of user concept map

The first step is to feed the pipeline with a conceptual model that we want to validate using the Algorithm 2.

Algorithm 2: Representing vertices and edges of user concept map

```

1 function initialize(mapFile.csv);
   Input : user concept map in CSV as mapFile.csv; // edge is a nX2 matrix
   Output: Initialization of user concept map into the pipeline
2 file ← csv.reader(mapFile); // CSV file reading using Python version
   3.7.1 csv.py module
3 for row ∈ file do
4   | G.add_edge(row[0],row[1]); // using NetworkX
5 end
6 return G

```

Here the conceptual model we are using is in the form of concept map as shown in Figure 3.5. Using NetworkX library in Python language, the map is initialized with its edges shared in the form of CSV file. The format of our CSV is shown in Figure 3.3.

From	To
age	exercise
intervention	exercise
income	exercise
income	obesity stigma
exercise	physical fitness
exercise	obesity
exercise	depression
obesity stigma	weight discrimination
weight discrimination	depression
stress	depression
stress	eating
stress	physical fitness
depression	antidepressant
obesity	weight discrimination
obesity	physical fitness
eating	obesity
antidepressant	eating

Figure 3.3: CSV format for concept map to be fed to the pipeline

3.3.3 Mapping derivationally related form onto concept nodes in lemma form

For each node of the user concept map (represented using NetworkX), map derivationally related form onto concept nodes in lemma form. This implies an expansion of all nodes name with their lemmatized derivationally related forms (using WordNet NLTK) because a **wordform** may not match the tweets (which were lemmatized using Stanford CoreNLP library [125]) content otherwise. Since retrieving tweets that include exactly a nodes name is too restrictive, derivationally related forms would be used. In WordNet, each synset contains one or more lemmas, which represent a specific sense of a specific word. Note that some relations are defined by WordNet only over Lemmas. The relations that are currently defined in this way are antonyms, **derivationally_related_forms** (see code below) and pertainyms. Antonyms are the words opposite in meaning to another. **derivationally_related_forms** are the words in different syntactic categories that have the same root form and are semantically related. Whereas pertainyms are relational adjectives, it can point to a noun or another pertainym.

```
from nltk.corpus import WordNet as wn
synset = wn.synsets ('study', 'n')[1]
lemma = synset.lemmas ()[0]
print (lemma.derivationally_related_forms ())
```

Algorithm 3 explains how to map derivationally related form onto concept vertices in lemma form.

Algorithm 3: Mapping derivationally related form onto concept vertices in lemma form

```

1 function getDerivationallyRelatedform( $V_i$ );
   Input : User concept map as  $G$ ; //  $G = (V, E)$  using function
           initialize(mapFile.csv)
   Output: der_rel_form[ $V_i$ ]
2 for  $V_i \in G$  do
   | tokens  $\leftarrow$  word.tokenize( $V_i$ ); // using nltk.tokenize package
3   for  $t \in$  tokens do
   | | der_rel_form[ $t$ ]  $\leftarrow$  getDerivationallyRelatedform[ $t$ ]; // using
   | | derivationally_related_forms relations from WordNet(NLTK))
4   end
5 end
6 return der_rel_form[ $V_i$ ]

```

3.3.4 Mapping relevant tweets onto concept nodes

Algorithm 4: Mapping relevant tweets onto concept vertices

```

1 function retrieveTweets( $V_i$ );
   Input : der_rel_form( $V_i$ )
           Twitter data as  $D$ 
   Output: relevantTweets[ $V_i$ ]
2 for  $k, v$  in der_rel_form( $V_i$ ) do
3   for keywords in  $v$  do
4   | for tweet in tweets do
5   | | if keywords in tweet then
6   | | | relevantTweets[ $k$ ]  $\leftarrow$  tweet; // mapping corresponding
   | | | vertices to the tweets that contain them using their
   | | | derivationally related form
7   | | end
8   | end
9   end
10 end
11 return relevantTweets[ $V_i$ ]

```

At this step of our proposed methodology (see Figure 3.2) using set of keywords fetched using `der_rel_form` (node) function, the pipeline finds the corresponding relevant tweets from the cleaned Twitter data and stores them in `relevantTweets` dictionary corresponding to user concept maps' nodes. The algorithm to map relevant tweets onto concept nodes is explained in Algorithm 4.

3.3.5 Extracting themes from mapped relevant tweets

Algorithm 5: To retrieve relevant themes using relevant tweets for each vertices of user concept map

```

1 function extractThemes( $V_i$ );
   Input : retrieveTweets( $V_i$ )
   Output: relevantThemes( $V_i$ )
2 for  $k,v$  in retrieveTweets( $V_i$ ) do
3   |  $themes[k] \leftarrow LdaMulticore(relevantTweets[k])$  ; // using gensim LDA
   | multicore model
4   |  $cleantthemes[k] \leftarrow themes[k]$  ; // using Google Cloud natural language
   | API to remove non-entities
5   |  $relevantThemes[k] \leftarrow cleantthemes[k]$  ; // using scikit-learn count
   | vectorizer and tf-idf transformer
6 end
7 return relevantThemes( $V_i$ )

```

In algorithm 5 for each set of relevant tweets we retrieve the relevant themes. This implies: Finding the prevalent themes with the help of gensim [126] Latent Dirichlet Accuracy (LDA) multicore model [127] which utilizes all CPU and GPU cores (explained in section Optimization) to parallelize and speed up model training, cleaning those themes that are already present in the set of derivationally related form of the node, Using Google Cloud Natural language API [128] to get rid of non-entities from the cleaned themes and Classifying the themes from most important to least important depending on their tf-idf weight which was calculated using scikit learn [129] [130] CountVectorizer and TfidfTransformer over the set of relevant tweets for each theme keyword.

3.3.6 Validating edges of user concept map

At this stage, the pipeline compares the theme corresponding to a node with derivationally related forms of all other nodes. If they are same, then create an edge consisting of a node having that theme keyword and the node having that theme keyword in its derivationally related form set of words. Finally, we confirm the edges if the edge created already exists in the user concept map. For later use, also suggest the remaining created edges for expansion. The algorithm for the last step of our methodology is explained in Algorithm 1 starting from step 6.

3.4 Experiment

3.4.1 Experimental set-up

The conceptual model in the form of concept map we are going to use to validate is shown in Figure 3.5 which was derived using the causal map shown in Figure 3.4.

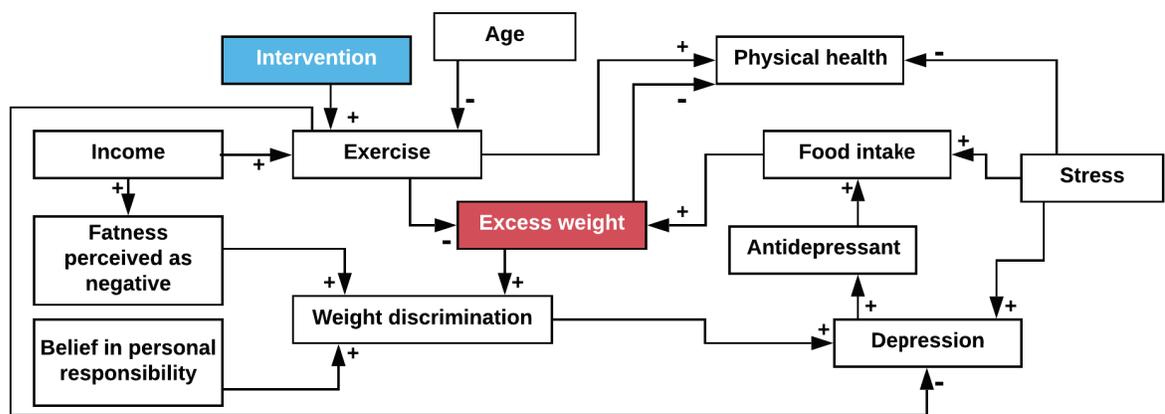


Figure 3.4: Causal Map - Obesity

This derivation was done by replacing following nodes: ‘Fatness perceived as negative’ with ‘obesity stigma’, ‘Physical health’ with ‘physical fitness’, ‘Excess weight’ with ‘obesity’ and ‘Food intake’ with ‘eating’, and removing the node ‘Belief in personal responsibility’. Concept names represented using phrases instead of words get more complicated to understand as well as people generally do not tweet these phrases, therefore for the sense of a clear concept map we used meaningful and single word representation of those phrases. The causal map in Figure 3.4 is a part of the conceptual model that was developed with the Provincial Health Services Authority (PHSA) of British Columbia to

explore the interrelationships involved in obesity and well-being. The model was presented in 2015 at the Canadian Obesity Summit [17] and tested with policymakers in 2016 [56].

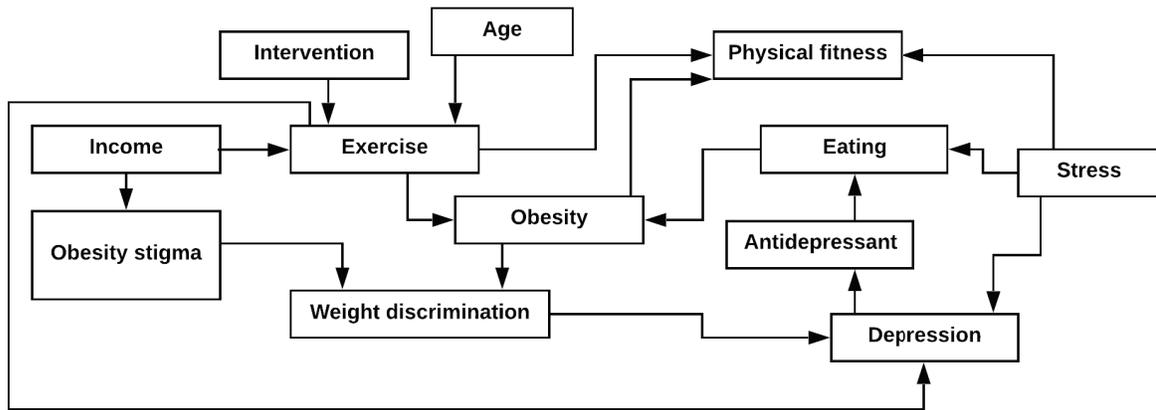


Figure 3.5: Concept Map - Obesity

The social media data, Twitter data in our case, was collected at the DataLab¹. The data for the concept ‘obesity was collected using the keywords shown in Table 3.1 consisting of 12 nodes and 17 edges. Note that the above-shared concept map is unweighted as the strength of evidence is used in later modelling stages going beyond the present scope. These keywords were further collected using the nodes of the map that we are going to use (see Figure 3.5) along with their respective derivationally related forms using WordNet [131] from Natural Language Toolkit (NLTK [132]).

¹DataLab url <http://www.datalab.science/>.

Table 3.1: Keywords used for topic ‘Obesity’ for Figure 3.5 (fetched using WordNet(NLTK) lemmatized derivationally related form)

Nodes of Map	List of related keywords using derivationally related form (Wordnet)
age	geezer, maturation, senescent, age
exercise	employ, exercise, utilise, workout, utilize, work_out, drill, exercising, exerciser, practice, use
intervention	interpose, intervention, treat, interfere, intervene, intercede, intervention
income	income
obesity stigma	obese, stigmatize, brand, stigma, mark, stigmatic, stain, fleshy, obesity stigma, obesity
physical fitness	physical, fitting, physicalness, fitness, fit, seaworthy, physical fitness, physicality, force, physics
obesity	obese, fleshy, obesity
depression	depression, slump, depress
weight discrimination	discriminate, weight discrimination, slant, weighty, discrimination, burden, weight, burden
stress	accentual, emphatic, stressor, tense, accentuation, strain, emphasis, accent, emphasizing, stress, accentuate
eating	eating, feeding, feed, eater, feeder, consumptive, exhaustion, consumable, corroding, depletion, eat, corrosion, rust, corrosive
antidepressant	antidepressant

Before selecting derivationally related form for retrieving tweets, we tested the validation results using synonyms as well. From the WordNet glossary [133]: Synonyms = synset+lemma, where synset is defined as a synonym set; a set of words that are interchangeable in some context without changing the truth value of the preposition in which they are embedded and lemma is defined as lower case ASCII text of word as found in the WordNet database index files. Lemma is usually the base form for a word or col-

location. Furthermore, derivationally related forms are the terms in different syntactic categories that have the same root form and are semantically related. In conclusion, not only did derivationally related form of the nodes of the map gave us meaningful single search keywords unlike synonyms (see Figure 3.6) but also 2% better validation rate than the keywords fetched using synonyms. The percentage of validation for synonyms using our previous works[103] Twitter data and conceptual model setup gave us 26 percent validation, whereas we received 56.5 percent validation from the derivationally-related form keywords.

Concept nodes	Synonyms	Derivationally related forms
Obesity	corpulency, obesity, fleshiness	obese, fleshy, obesity
Stress	emphasize, accent, focus, emphasis, emphasise, stress, try, accentuate, punctuate, tension, tenseness, strain	accent, emphasizing, emphasis, tense, stress, accentuate, emphatic, accentuation, stressor, accentual, strain
Depression	imprint, economic_crisis, clinical_depression, depressive_disorder, great_depression, depression, low, natural_depression, slump, impression	depress, depression, slump
Eating	consume, use_up, exhaust, feeding, run_through, eat, wipe_out, corrode, eat_up, eat_on, feed, deplete, eating, rust	eater, corrosion, feeder, feeding, eat, corrosive, feed, rust, eating, corroding, exhaustion
Coronary	coronary_thrombosis, coronary	corona, coronary

Figure 3.6: Comparison between synonyms and derivationally related form keywords fetched using WordNet NLTK for concept nodes defined in Figure 2.2

In total, 5,000,001 raw tweets were collected in this process from Feb. 26, 2019 to Mar. 2,

2019 for concept ‘obesity. For this research, we have used English language tweets. This data was used to develop and validate concept maps by mining Twitter data. In order to work with human language data, Natural Language ToolKit (NLTK) provides us with an interface like WordNet that we are using to get the `derivationally_related_forms` relations which are defined only over lemmas. For retrieving lemmatized tweets from Twitter data, we have used Stanford’s coreNLP library which provides us with fast and robust annotators to manipulate and analyze text. We have used the following annotators in our script from this library: `tokenization`, `sentence splitting`, POS tagging, and `lemmatization`.

3.5 Implementation

Table 3.2: Versions of libraries used

Libraries	Versions	Used for
stanford-corenlp	3.9.2	Lemmatization
networkx	2.2	Conceptual model (accessing node labels and edges)
nlTK	3.4	Wordnet (derivationally related forms)
gensim	3.6.0	Parallel, multi-core Latent Dirichlet Allocation (LDA) model for big data
google-cloud-language API	1.1.1	Entity identification
scikit-learn	0.20.1	Sorting words by tf-idf

For the development of the complete framework, we relied on the technologies listed in Table 3.2. NetworkX was used to create, manipulate and analyse the structure of complex networks like our conceptual maps. It has the ability to use and work with large data-sets and create maps. We used it for visual representation and initialization of our maps. We replaced Latent Dirichlet Allocation (LDA) model with LDA Multicore model is also known as Online LDA model (from Gensim library) which uses all the CPU or GPU cores (see section Optimization) assigned to it. This model does parallelization that uses multiprocessing to speed up the topic modelling process. Since we were using

millions of Twitter data-set, this model gave us faster results on a High Performance Computational (HPC) system with 3 cores assigned to each script that was used to fetch the validated map. Furthermore, on a GPU system, this model ran 2 times faster than the older model running on CPUs(see section Optimization). In our LDA model, the following parameters (see table B.1) were used to determine relevant topics:

```
LDA = gensim.models.LdaMulticore (corpus=corpus, id2word=dictionary,
    num_topics=n_topics, chunksize=10000, passes=20, eval_every = None,
    workers=3, dtype=np.float64, iterations = 400)
```

- **corpus**(iterable of list of (int, float), scipy.sparse.csc, optional) = Stream of document vectors or sparse matrix of shape (num_terms, num_documents). If not given, the model is left untrained.
- **id2word**(dict of (int, str), gensim.corpora.dictionary.Dictionary) = Mapping from word IDs to words. It is used to determine the vocabulary size, as well as for debugging and topic printing.
- **num_topics**(int, optional) = The number of requested latent topics to be extracted from the training corpus.
- **chunksize**(int, optional) = Number of documents to be used in each training chunk.
- **passes**(int, optional) = Number of passes through the corpus during training.
- **eval_every**(int, optional) = Log perplexity is estimated every that many updates. Setting this to one slows down training by 2 times.
- **workers**(int, optional) = Number of workers processes to be used for parallelization.
- **dtype**(numpy.float16, numpy.float32, numpy.float64, optional) = Data-type to use during calculations inside model. All inputs are also converted.
- **iterations**(int, optional) = Maximum number of iterations through the corpus when inferring the topic distribution of a corpus.

The documents, in our case set of tweets are represented as strings. We used a document representation called bag-of-words to convert documents into vectors. In this each document is represented by one vector in which each vector element represents a question-answer pair, in the style of: "How many times does the word **obesity** appear in the document? Twelve times." Representing the questions by their (integer) ids alone is beneficial. A **dictionary** is called the mapping between the questions and ids. We assigned a unique integer id to all words that appear in the corpus with the `gensim.corpora.dictionary.Dictionary` class. This sweeps through the texts, gathering relevant word numbers and statistics. We used the function `doc2bow()` to convert tokenized documents into vectors that simply counts the number of occurrences of each distinct word, converts the word to its integer word id and returns the result as a sparse vector. Now, we have arrived at the corpus of vectors. We have millions of documents in the corpus, storing all of them in RAM wont do. Gensim only requires a corpus to return one vector of a document at a time. So we parse our input to get a clean list of tokens in each document, then convert the tokens to their ids through a dictionary and deliver the resulting sparse vector. Now the corpus is much more memory-friendly because at most one vector resides at a time in RAM. Corpus is now an object which is provided as inout for parameter `corpus`. The parameter `id2word` uses the dictionary we created above as input. `num_topics` parameter takes in input our global parameter `n_topics`. The parameter `chunksize` is set to 10000 number of documents to consider at once since it affects the memory consumption. `Passes` set to 20 which defines how many times the algorithm is supposed to pass over the whole corpus for training the model. To make sure that by the final passes, most of the documents have converged Parameter `iterations` was set to 400. Since parameter `eval_every` slows down the process if set to 1, so we chose to input 'none' for faster computations. Paramter `workers` was set to 3 considering that it was measured on i7 server with 4 physical cores[127], so the optimal `workers=3`, one less than the number of cores. For precision we used `dtype` used for calculations as `numpy.float64`(Double precision float). Also, parameter α which is the per-document topic distributions and β which is the per-topic word distribution were tested before finalizing these model constraints for topic detection. Previously we used them to improve results, but it failed for our Twitter corpus which is random documents (tweets). In our case document is equivalent to one tweet unlike other models [134] which is one whole article. α , β parameters specify prior beliefs about topic sparsity/uniformity in the documents which in our data couldn't be predicted due to the limitation of random opinions shared on Twitter. Later, for entity analysis to get rid of non-entities

from our fetched themes we have integrated Googles Cloud Natural Language API in our script as well. Furthermore to fetch important themes Scikit-learn's countvectorizer and tf-idf transformer were used as a weighing factor. At this step our third global paramter `max_df` which is the tf-idf weight threshold is set between range 2 to 9 since the tf-idf weights of a word in our Twitter data for any value of other global parameters `n_topics` and `n_words` was between 0 to 10.

3.6 Discussion

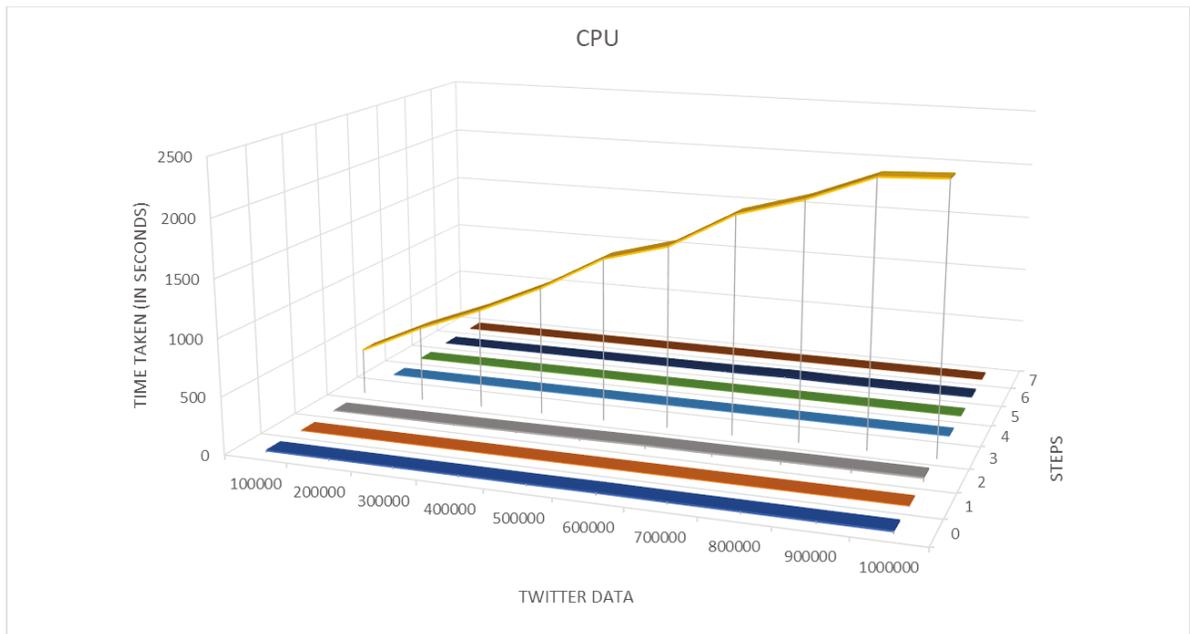
The experimental setup explained in section Experiment was used to run the following computations. Seven main steps of our pipeline were tested for the time taken to execute them. Time taken for execution was calculated for module data upload, using derivationally related form from WordNet (NLTK), retrieving relevant tweets, extracting relevant themes, cleaning repetitive themes, removing non entities using Google Cloud Natural Language API, filtering the most important themes using count vectorizer and tf-idf transformer weights and at last validating the connection in our user concept map (see Figure 3.5). This time was calculated using the below python library which gives us CPU and wall clock times for a particular function running in a Jupyter notebook cell:

```
%time functionname()
```

Now these steps were run on the Twitter data mentioned in the experiment section 3.3 with the increments of 100,000 tweets from 100,000 to 1,000,000 (out of 5,000,001 initial tweets). These chunks of tweets were collected one after the other not randomly. Not only did we calculate the time taken at each step relative to the number of tweets, but also the number of confirmed edges out of 17 nodes from our concept map (explained previously in section Experimental set-up). This optimization was tested on Intel CORE i7 8th Gen CPU as well as Nvidia 410.78 GPU. In Figure 3.7a, the table shows us the time taken by steps 0 to 7 in seconds when executed on a CPU machine with an increment of 100,000 tweets for validating connections from user concept map (see Figure 3.5). In Figure 3.7b the graphical representation for the time taken by each step (0-7) for incremental Twitter data (100,000-1,000,000) on CPU shows us that with the increase in the number of tweets, the time taken to execute: step 2 = fetch relevant tweets and step 3 = LDA increases.

CPU Results (intel CORE i7 8th Gen)										
for n=50/50/9 new 5m data/17 edges map	data upload	lemmatized der_rel_form	relevant tweets	relevant themes	clean themes	non-entity removal	tf-idf filtering	validate connection	confirmed %	out of 17
Data/Steps	0	1	2	3	4	5	6	7		
100000	14.718	1.452	3.205	406.712	0	0.001	0.938	0.019	64.7058824	11
200000	14.913	1.471	6.047	667.461	0	0.001	1.943	0.016	82.3529412	14
300000	14.741	1.498	8.927	887.524	0	0.002	2.394	0.018	76.4705882	13
400000	15.548	1.536	12.236	1139.844	0	0.002	2.893	0.019	76.4705882	13
500000	15.475	1.596	17.345	1450.272	0	0.003	3.99	0.026	82.3529412	14
600000	18.428	1.993	23.283	1605.693	0.001	0.004	5.389	0.027	82.3529412	14
700000	16.848	1.939	23.508	1923.215	0.001	0.004	13.052	0.06	76.4705882	13
800000	16.852	1.777	27.931	2090.032	0.001	0.005	5.843	0.02	76.4705882	13
900000	16.317	1.807	29.338	2313.806	0.002	0.005	7.677	0.026	82.3529412	14
1000000	16.952	1.661	30.416	2338.468	0.001	0.006	7.691	0.025	76.4705882	13

(a) Time taken



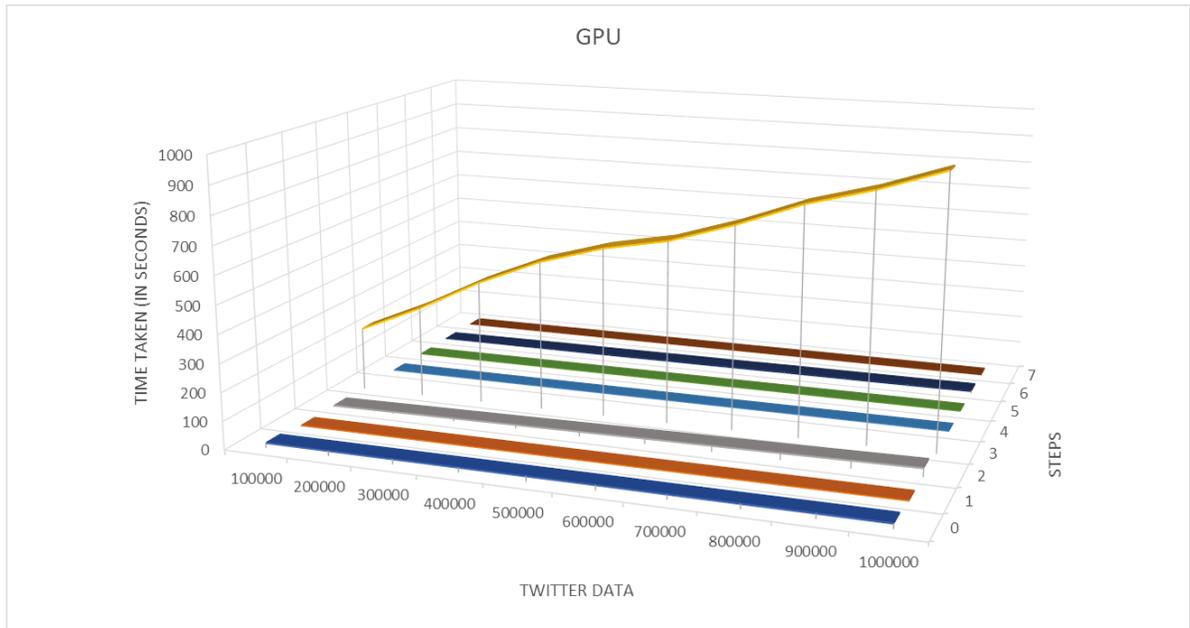
(b) Graphical representation

Figure 3.7: Computation results on CPU (Intel CORE i7 8th Gen) - (a) Time taken by 7 steps and number of confirmed edges for data from (100,000 - 5,000,000) tweets with increment of 100,000 tweets on CPU and (b) Graphical representation for time taken by each step (0-7) for incremental Twitter data (100,000-1,000,000) on CPU

In Figure 3.8a, the table shows us the time taken by steps 0 to 7 in seconds when executed on a GPU machine with increment of 100,000 tweets for validating connections from user concept map (see Figure 3.5). In Figure 3.8b the graphical representation for time taken by each step (0-7) for incremental Twitter data (100,000-1,000,000) on GPU shows us that with increase in number of tweets, the time taken to execute: step 2 = fetch relevant tweets and step 3 = LDA increases.

GPU Results (nvidia 410.78)										
for n=50/50/9 new 5m data/17 edges map	data upload	lemmatized der_rel_form	relevant tweets	relevant themes	clean themes	non-entity removal	tf-idf filtering	validate connection	confirmed %	out of 17
Data/Steps	0	1	2	3	4	5	6	7		
100000	15.571	1.811	2.835	228.153	0.001	0.001	0.948	0.014	76.4705882	13
200000	15.488	1.82	5.776	327.819	0.001	0.001	1.71	0.015	76.4705882	13
300000	15.52	1.845	8.465	443.583	0.001	0.001	2.875	0.019	76.4705882	13
400000	15.826	1.886	11.42	539.245	0.001	0.001	2.959	0.015	76.4705882	13
500000	15.787	1.89	14.218	606.355	0.001	0.002	3.685	0.022	82.3529412	14
600000	16.074	1.95	17.378	650.91	0.001	0.002	3.984	0.016	88.2352941	15
700000	15.692	1.91	19.643	724.387	0.001	0.002	5.033	0.017	76.4705882	13
800000	15.687	1.927	22.494	812.176	0.001	0.002	5.32	0.017	88.2352941	15
900000	15.892	1.967	25.492	875.904	0.001	0.002	5.739	0.018	76.4705882	13
1000000	15.531	1.971	28.645	955.215	0.001	0.004	6.807	0.026	88.2352941	15

(a) Time taken



(b) Graphical representation

Figure 3.8: Computation results on GPU (nvidia 410.78) - (a) Time taken by 7 steps and number of confirmed edges for data from (100,000 - 5,000,000) tweets with increment of 100,000 tweets on GPU and (b) Graphical representation for time taken by each step (0-7) for incremental Twitter data (100,000-1,000,000) on GPU

3.7 Conclusion

The pipeline to validate complex conceptual models by mining Twitter data can be used as a generic methodology for validating conceptual models using citizen science approach. Our dependency is on the type of Twitter data-set that we use to validate these conceptual models, which is accomplished by using WordNet derivationally related form to fetch the related keywords of the user concept map and is used to fetch tweets

from the Twitter API on which we validate these models. Also, our proposed optimized pipeline using GPU gave us two times faster results than our computations on CPU. Also our user-defined global parameters `n_topics`, `n_words` and `max_df` were tested for combination of values. These values ranged from 10 to 100 with increment of 10 for `n_topics` and `n_words` whereas 1 to 10 with increment of 1 for `max_df`.

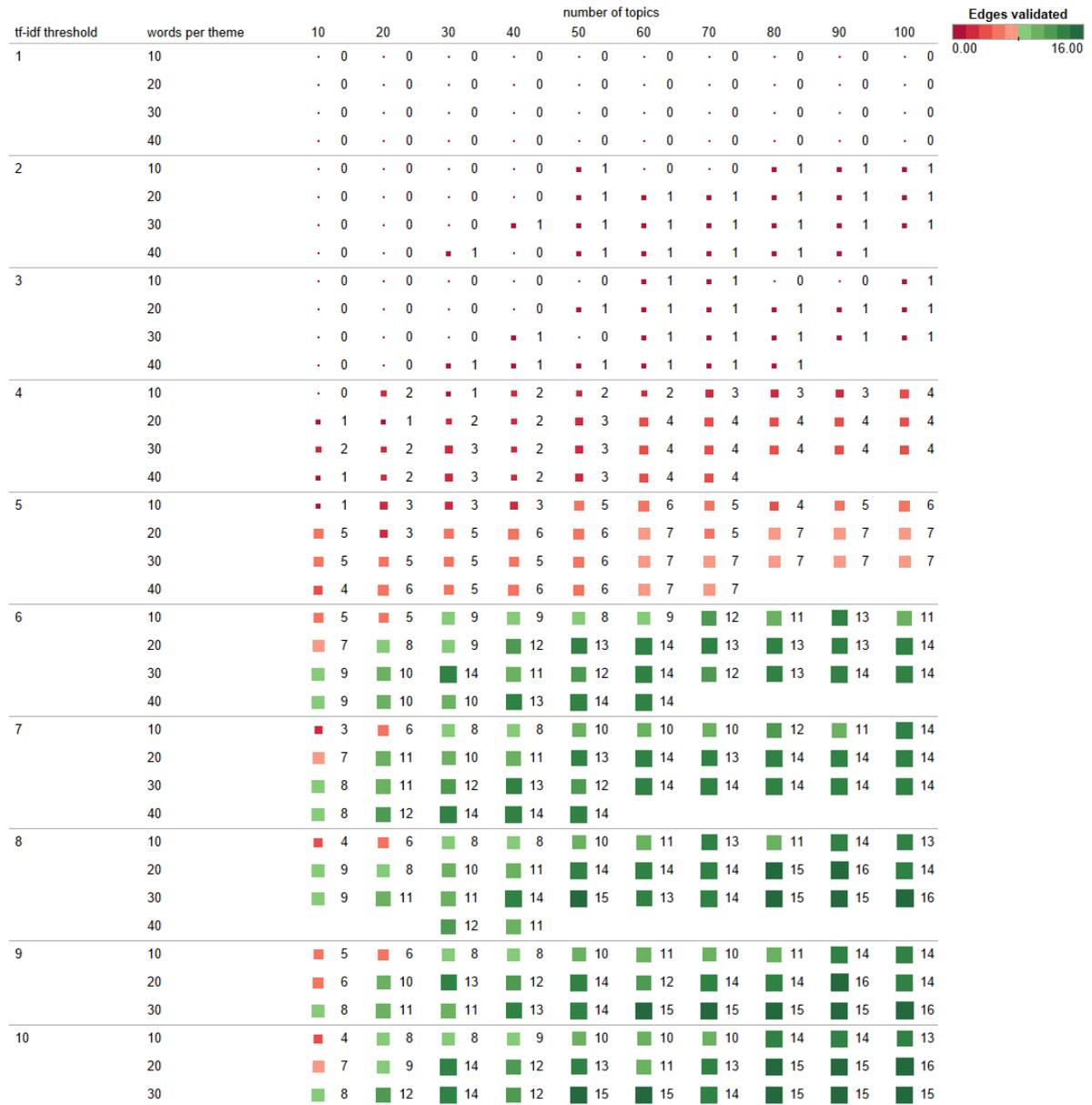


Figure 3.9: Number of edges confirmed (out of 17 in Figure 3.5) for each combination of parameter values using Twitter data

The results (see Figure 3.11) shows us that with increase in `n_topics`, `n_words` and

`max.df` values, the validation percentage also increases. Here, we used the experimental set up explained previously in section Experiment. The maximum number of validated edges of user concept map was 16 out of 17 which 94.12%. This concludes that with a large number of topics and words Twitter data provides a good validation percentage with the most important words using maximum tf-idf weighing.

Chapter 4

Extending conceptual models using Granger causality by mining literature and Twitter conversations

4.1	Abstract	48
4.2	Introduction	49
4.3	Methodology	50
	4.3.1 Data Collection	50
	4.3.2 Data preprocessing	53
	4.3.3 Test for Causality	53
4.4	Results	54
4.5	Conclusion	59
	4.5.1 Limitations	59
	4.5.2 Future work	60

4.1 Abstract

This chapter explores the possibility of Granger causality between the edges of a conceptual map, measure in logarithms of real absolute frequency of the edges of a concept map in 20 academic papers/journals/articles since January 01, 2019. By using the results from this test, we can confirm and extend the edges of a concept map and generate

a causal map out of it. A causal map has directed and labeled edges. So far approaches can only confirm or suggest undirected, unlabeled edges. That is, we can confirm that two constructs are associated (one comes up as key theme when the other is evoked), but we cannot identify which one causes the other, and what type of causation. By using Granger causality tests, if the hypothesis that V_1 does not Granger-causes V_2 is Rejected, we can confirm the causality between two concept vertices (V_i) and extend to the reverse causality if the vice-versa is Rejected.

To find causality, we need to follow a time series. This time series is generated using the absolute frequency of the vertices that create an edge over a period of one month since January 01, 2019 using literature. Granger causality is being used once we have the time series. Later we discuss the use of Twitter conversations instead to evoke temporal precedence which determines the strength of a cause and effect relationship. This section will reflect the relevance of Twitter conversations as compared to literature to extend the conceptual model and generate strength of evidence associated with the model. We highlight the basic ideas they are based on, provide a comparative analysis, and point to some fundamental issues.

4.2 Introduction

After validating a conceptual model, one questions the extent and direction of validation. Previously there was no evidence of strength provided for associations between concepts like ‘Obesity’ and ‘Stress’ for health-related conceptual models [See Chapter 3]. Establishing an association does not necessarily mean that the effect ‘Obesity’ is the cause of the outcome ‘Stress’ or vice-versa. Most definitions of ”cause” include the idea that it is something that has an effect or a consequence. Establishing a valid association between ‘Obesity’ and ‘Stress’ is a necessary first step that was accomplished before using our proposed pipeline [See Figure 3.2] . Whether this relationship is causal, is the more complicated and often debatable question. There are no universal rules to determine the causality of a relationship.

In 1890 Robert Koch [135] proposed specific criteria that should be met before concluding that a disease was caused by a particular bacterium. These became known as Koch’s Postulates which established standard criteria for concluding the cause of infectious disease, but the criteria had some limitations concerning non-infectious diseases.

Simultaneously in the past, efforts were made to account for the occurrence of disease outcomes by applying Hill's criteria for a judgment of causality [136] which was later argued by Charlton [137] to be not used to assess causation. This was because of Hill's pre-defined criteria just like Robert Koch's postulates which diminishes the validity of causal inferences. On the contrary, Wynder [138] argued that criteria should be used more often than they are to assess causation. According to him, these criteria reduce the tendency for investigators to make causal conclusions merely on the basis of their own published results of the association in question and thereby increase the validity of causation. A recent study [139] of these practices reveals errors in two claims made by Charlton and Wynder. Research on causal inference methodology [140, 141] is encouraged, including research on the underlying theory, methodology, and additional systematic descriptions of how causal inference is practiced [142, 143].

In recent years, Granger causality [144, 145] has emerged as a leading technique for inferring directions directly from data of neural interactions and information flow [146]. Our research focuses on using this technique to recognize the importance of temporal order for causal relationship inference which was recognized by Wiener in 1956 [147] and then formulated by Granger in terms of autoregressive (AR) models of time series in 1969 [144]. Unlike stimulation using predefined criteria, the Wiener Granger method does not require direct intervention in complex systems. Rather, it relies on the evaluation of causal statistical influence between simultaneously recorded time series data for concepts of our complex system [See Figure 4.1]. Causality in the WienerGranger[144, 147] sense is based on the statistical predictability of one time series that derives from knowledge of one or more others. Our research proposes a new method to get the strength of evidence i.e. causality for each concept defined by our conceptual model by mining Literature and then compares it with Twitter conversations mining.

4.3 Methodology

4.3.1 Data Collection

To discuss the causality of concept maps generated by mining literature and Twitter as knowledge base, our data collection was divided into two parts - one using literature related PDFs and the other using Twitter conversations. The concept map used for this research analysis is shown in Figure 4.2.

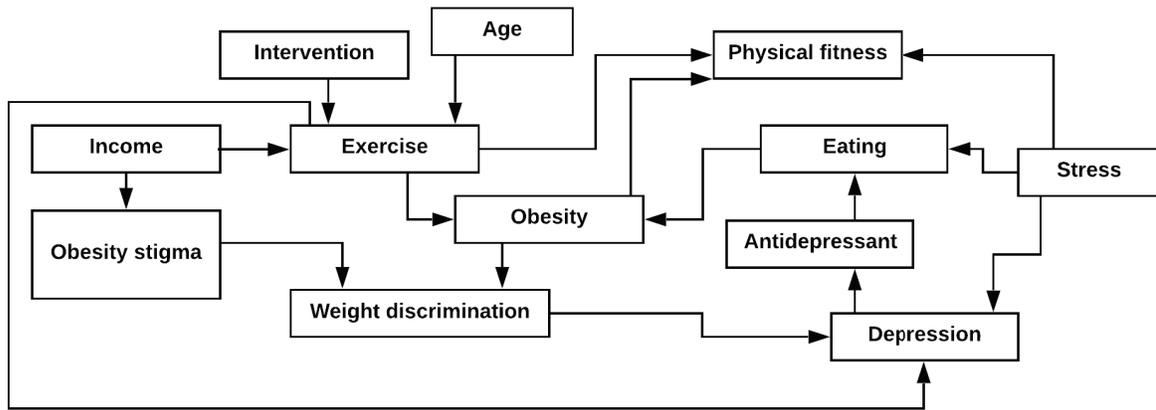


Figure 4.1: Conceptual map

For the literature mining, the data in PDF format was collected using Google Scholar[148], The Journal of Medical Internet Research[149], BioMed Central[150], PLOS[151] and IEEE[152] websites. For each concept edge, we collected 20 PDFs. Each PDF had in its title one of the vertices of that edge [See Figure 4.2a]. The time-frame used to collect these PDFs was from January 01, 2019 till January 31, 2019.

On the other hand for Twitter conversation mining, we manually collected conversations around the concept edges (‘obesity’, ‘stress’) and (‘obesity’, ‘education’) only as a proof of concept [See Figure 4.2b]. To assess the temporality of tweets we need to follow conversations (i.e. when users answer each other and the conversation brings new constructs). The process for collecting Twitter conversations was a challenge in this research due to the Twitters policy which discusses the affordability of the Twitter API for ‘replies’ since they create conversations. Tweets retrieved from the Twitter API are in JSON, a simple structured text format. Twitter provides documentation on the complete set of fields in a tweet [153]. But there is nothing to indicate that a tweet has a reply. Instead to find tweet to which the reply tweet is a reply, Twitter provides the following relevant fields in reply tweet `in_reply_to_status_id`, `in_reply_to_status_id_str`, `in_reply_to_screen_name`, `in_reply_to_user_id`, `in_reply_to_user_id_str`. From the names of these fields, one can easily conclude what they imply. The metadata of a tweet provides the field `in_reply_to_status_id` which can be used to follow a chain of replies backward from the reply tweet to the replied to tweet, but not vice versa, i.e., from the replied to tweet to the reply tweet. This is where the challenge arises due to its comprehensive selection of replied to tweets id i.e. on what relevance basis do we select

these conversations which are beyond the present scope.



Nutrition 23 (2007) 887–894

NUTRITION

www.elsevier.com/locate/nut

Review

Relationship between stress, eating behavior, and obesity

Susan J. Torres, M.Nutr.Diet.*, and Caryl A. Nowson, Ph.D.

Centre for Physical Activity and Nutrition, School of Exercise and Nutrition Sciences, Deakin University, Burwood, Victoria, Australia

Manuscript received March 3, 2007; accepted August 13, 2007.

Abstract

Stress is thought to influence human eating behavior and has been examined in animal and human studies. Our understanding of the stress-eating relation is confounded by limitations inherent in the study designs; however, we can make some tentative conclusions that support the notion that stress can influence eating patterns in humans. Stress appears to alter overall food intake in two ways, resulting in under- or overeating, which may be influenced by stressor severity. Chronic life stress seems to be associated with a greater preference for energy- and nutrient-dense foods, namely those that are high in sugar and fat. Evidence from longitudinal studies suggests that chronic life stress may be causally linked to weight gain, with a greater effect seen in men. Stress-induced eating may be one factor contributing to the development of obesity. Future studies that measure biological markers of stress will assist our understanding of the physiologic mechanism underlying the stress-eating relation and how stress might be linked to neurotransmitters and hormones that control appetite. © 2007 Elsevier Inc. All rights reserved.

Keywords:

Stress; Eating behavior; Sucrose; Fat; Rat; Weight gain; Obesity

(a) Literature



(b) Twitter conversation

Figure 4.2: Data collection for Granger causality - (a) Literature - for concept edge ('obesity', 'stress') PDF 'Relationship between stress, eating behavior, and obesity'[154] with 'obesity' and 'stress' in PDF title and (b) Twitter conversation - for concept edge ('obesity', 'stress')

4.3.2 Data preprocessing

Initially, all PDFs were converted to raw text using XpdfReader's[155] command line tool pdftotext. While collecting the data in PDF format, we have extracted the references, images, tables as well. Each text file is divided into sentences using NLTK sentence tokenizer (`nltk.tokenize` package). Finally, this text file is passed through our proposed pre-processing techniques [See Figure 2.5] as explained previously in Chapter 2 section 2.5.1. Simultaneously, the same steps were reproduced for pre-processing Twitter conversations in the form of tweets.

4.3.3 Test for Causality

To investigate causality between two concepts in a time series we are going to use Granger causality tests from `statsmodels` Python module[156]. This module provides classes and functions for the estimation of many different statistical models, as well as for conducting statistical tests, and statistical data exploration. We used the following function from the module which computes four tests for Granger non-causality of two time series.

```
statsmodels.tsa.stattools.grangercausalitytests(x, maxlag, addconst=True,
        verbose=True)
```

It returns all test results in the form of a dictionary where keys are the number of lags. For each lag the values are a tuple, with the first element a dictionary with teststatistic, pvalues, degrees of freedom, and the second element are the OLS estimation results for the restricted model, the unrestricted model, and the restriction (contrast) matrix for the parameter `f_test` [156]. This method is a probabilistic account of causality; it uses empirical data sets to find patterns of correlation.

The Null hypothesis for `grangercausalitytests` is that the time series for vertex V_2 does not Granger cause the time series for vertex V_1 . The time series for each vertex is derived by calculating the absolute frequency of each concept name V_i in their respective data source over time. Using literature, the absolute frequency was calculated by the summation of the number of times the concept name along with its `derivationally_related_form` occurs in corresponding 20 PDFs collected over a time period of one month for all vertices. On the other hand, using Twitter conversations, the absolute frequency was calculated by the summation of the number of times the

concept name along with its `derivationally_related_form` occurs in the `replied_to` tweets for one vertex and `replied_by` tweets for another vertex over a period of one week.

We then reject the null hypothesis that V_2 does not Granger cause V_1 if the pvalues are below a desired size of the test. We chose 5% level of significance i.e. pvalue equivalent to 0.05. The level of significance meaning the probability of rejecting the null hypothesis when it is true which can be either 5%, 1% or 0.1% depending on the critical region. Considering we have a time series of 20 PDFs and subsequently tweets, pvalue of 0.05 gave us statistically significant results which considers less than 1 in 20 chance of being wrong. Unlike pvalue of 0.001 which provides statistically high significant results but by considering less than 1 in 1000 chance of being wrong. The null hypothesis for all four tests is that the coefficients corresponding to past values of the second time series are zero. Parameters `params_ftest`, `ssr_ftest` are based on F distribution whereas, `ssr_chi2test`, `lrtest` are based on chi-square distribution. F distributions are used in the analysis of variance i.e. the quality of being different. On the other hand, chi-square distributions provides goodness of fit i.e. how likely is an observed distribution is due to chance.

4.4 Results

The maximum lag achieved using `grangercausalitytests` was 7 for the literature dataset due to 20 PDFs over a period of one month. While testing when we increased the number of PDFs to 25 or 30, the maximum lag value also increased. The parameters `params_ftest`, `ssr_ftest`, `ssr_chi2test` and `lrtest` were generated [See Figure 4.3] for each hypothesis that V_2 does not Granger cause V_1 to confirm causality as well as V_1 does not Granger cause V_2 to extend reverse causality. This process was done for all Lag values ranging from 1 to 7 for the purpose of selecting the Lag value to finally generate a causal map [See Figure 4.5]. Figure 4.4 gives the overall computational results of Lags ranging from 1 to 7. We analyzed that Lag equivalent to 7 and test results from parameter `ssr_chi2test` reported 76.47% edge confirmation i.e. 13 out of 17 edges rejected the null hypothesis that V_2 does not Granger cause V_1 . 88.24% edges were extended in terms of reverse causality i.e. 15 out of 17 edges rejected the reverse null hypothesis that V_1 does not Granger cause V_2 .

Null hypothesis	no_of_doc	max_lag	lag	ssr based F test	p-value	ssr based chi2 test	p-value	likelihood ratio test	p-value	parameter F test	p-value
AGE does not Granger Cause EXERCISE	25	7	1	3.4118	0.0789	3.8992	0.0483	3.6131	0.0573	3.4118	0.0789
EXERCISE does not Granger Cause AGE	25	7	1	3.0991	0.0929	3.5419	0.0598	3.3037	0.0691	3.0991	0.0929
STRESS does not Granger Cause DEPRESSION	25	7	1	0.437	0.5157	0.4995	0.4797	0.4943	0.482	0.437	0.5157
DEPRESSION does not Granger Cause STRESS	25	7	1	0.0095	0.9232	0.0109	0.9169	0.0109	0.9169	0.0095	0.9232
STRESS does not Granger Cause EATING	25	7	1	0.0807	0.7792	0.0922	0.7614	0.092	0.7616	0.0807	0.7792
EATING does not Granger Cause STRESS	25	7	1	0.0611	0.8072	0.0698	0.7916	0.0697	0.7917	0.0611	0.8072
EXERCISE does not Granger Cause PHYSICAL FITNESS	25	7	1	0.7572	0.394	0.8654	0.3522	0.8502	0.3565	0.7572	0.394
PHYSICAL FITNESS does not Granger Cause EXERCISE	25	7	1	0.7757	0.3884	0.8865	0.3464	0.8705	0.3508	0.7757	0.3884
EXERCISE does not Granger Cause OBESITY	25	7	1	0.0721	0.7909	0.0824	0.774	0.0823	0.7742	0.0721	0.7909
OBESITY does not Granger Cause EXERCISE	25	7	1	2.5802	0.1231	2.9488	0.0859	2.7813	0.0954	2.5802	0.1231
EXERCISE does not Granger Cause DEPRESSION	25	7	1	7.9767	0.0102	9.1163	0.0025	7.7273	0.0054	7.9767	0.0102
DEPRESSION does not Granger Cause EXERCISE	25	7	1	1.7258	0.2031	1.9724	0.1602	1.8955	0.1686	1.7258	0.2031
OBESITY STIGMA does not Granger Cause WEIGHT DISCRIMINATION	25	7	1	1.3111	0.2651	1.4984	0.2209	1.4535	0.228	1.3111	0.2651
WEIGHT DISCRIMINATION does not Granger Cause OBESITY STIGMA	25	7	1	1.2772	0.2712	1.4596	0.227	1.417	0.2339	1.2772	0.2712
WEIGHT DISCRIMINATION does not Granger Cause DEPRESSION	25	7	1	0.5799	0.4548	0.6628	0.4156	0.6538	0.4188	0.5799	0.4548
DEPRESSION does not Granger Cause WEIGHT DISCRIMINATION	25	7	1	0.4117	0.528	0.4705	0.4927	0.466	0.4948	0.4117	0.528
STRESS does not Granger Cause PHYSICAL FITNESS	25	7	1	0.0051	0.9438	0.0058	0.9392	0.0058	0.9392	0.0051	0.9438
PHYSICAL FITNESS does not Granger Cause STRESS	25	7	1	0.1493	0.7031	0.1706	0.6795	0.17	0.6801	0.1493	0.7031
DEPRESSION does not Granger Cause ANTIDEPRESSANT	25	7	1	12.1862	0.0002	13.9271	0.0002	10.9827	0.0009	12.1862	0.0002
ANTIDEPRESSANT does not Granger Cause DEPRESSION	25	7	1	0.5643	0.4609	0.6449	0.4219	0.6364	0.425	0.5643	0.4609
OBESITY does not Granger Cause WEIGHT DISCRIMINATION	25	7	1	0.2282	0.6378	0.2608	0.6096	0.2594	0.6105	0.2282	0.6378
WEIGHT DISCRIMINATION does not Granger Cause OBESITY	25	7	1	1.6085	0.2186	1.8383	0.1751	1.7713	0.1832	1.6085	0.2186
OBESITY does not Granger Cause PHYSICAL FITNESS	25	7	1	1.953	0.1769	2.232	0.1352	2.1342	0.144	1.953	0.1769
PHYSICAL FITNESS does not Granger Cause OBESITY	25	7	1	0.4883	0.4924	0.558	0.4551	0.5516	0.4577	0.4883	0.4924
EATING does not Granger Cause OBESITY	25	7	1	1.8047	0.1935	2.0626	0.151	1.9787	0.1595	1.8047	0.1935
OBESITY does not Granger Cause EATING	25	7	1	2.2019	0.1527	2.5164	0.1127	2.3931	0.1219	2.2019	0.1527
ANTIDEPRESSANT does not Granger Cause EATING	25	7	1	0.4996	0.4875	0.571	0.4499	0.5643	0.4525	0.4996	0.4875
EATING does not Granger Cause ANTIDEPRESSANT	25	7	1	0.1686	0.6855	0.1927	0.6607	0.1919	0.6613	0.1686	0.6855
INTERVENTION does not Granger Cause EXERCISE	25	7	1	0.0056	0.9409	0.0064	0.936	0.0064	0.936	0.0056	0.9409
EXERCISE does not Granger Cause INTERVENTION	25	7	1	0.0039	0.9509	0.0044	0.9469	0.0044	0.9469	0.0039	0.9509
INCOME does not Granger Cause EXERCISE	25	7	1	0.0003	0.986	0.0004	0.9848	0.0004	0.9848	0.0003	0.986
EXERCISE does not Granger Cause INCOME	25	7	1	0.1408	0.7112	0.1609	0.6883	0.1604	0.6888	0.1408	0.7112
INCOME does not Granger Cause OBESITY STIGMA	25	7	1	1.969	0.1752	2.2503	0.1336	2.1509	0.1425	1.969	0.1752
OBESITY STIGMA does not Granger Cause INCOME	25	7	1	18.4858	0.0003	21.1266	0	15.154	0.0001	18.4858	0.0003

Figure 4.3: Parameter tests for each hypothesis

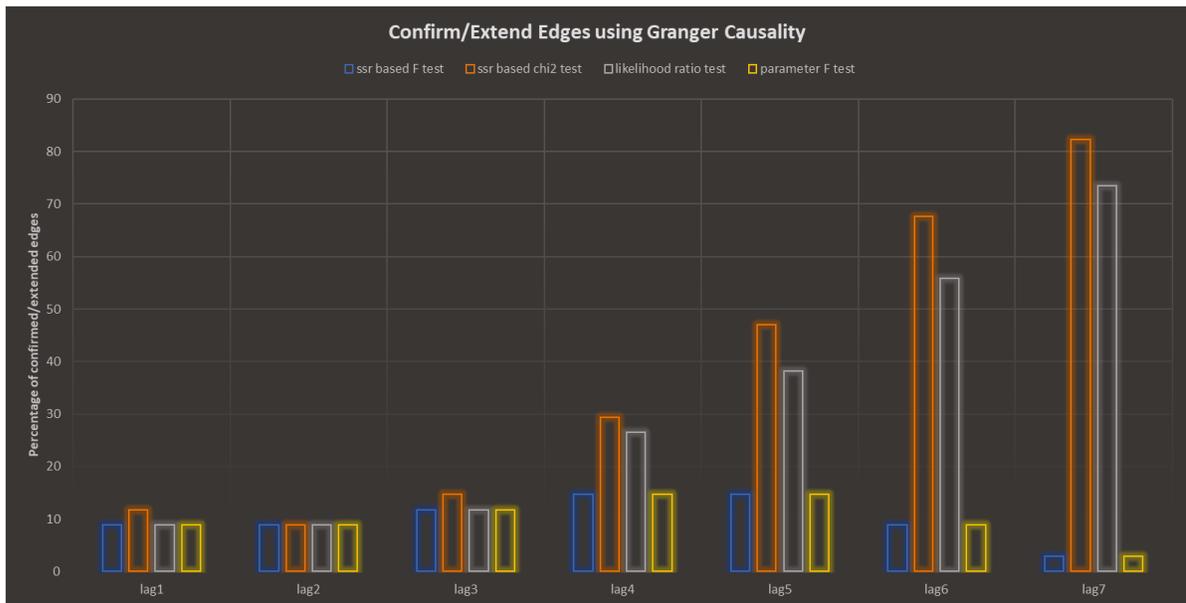


Figure 4.4: Optimal Lag length selection

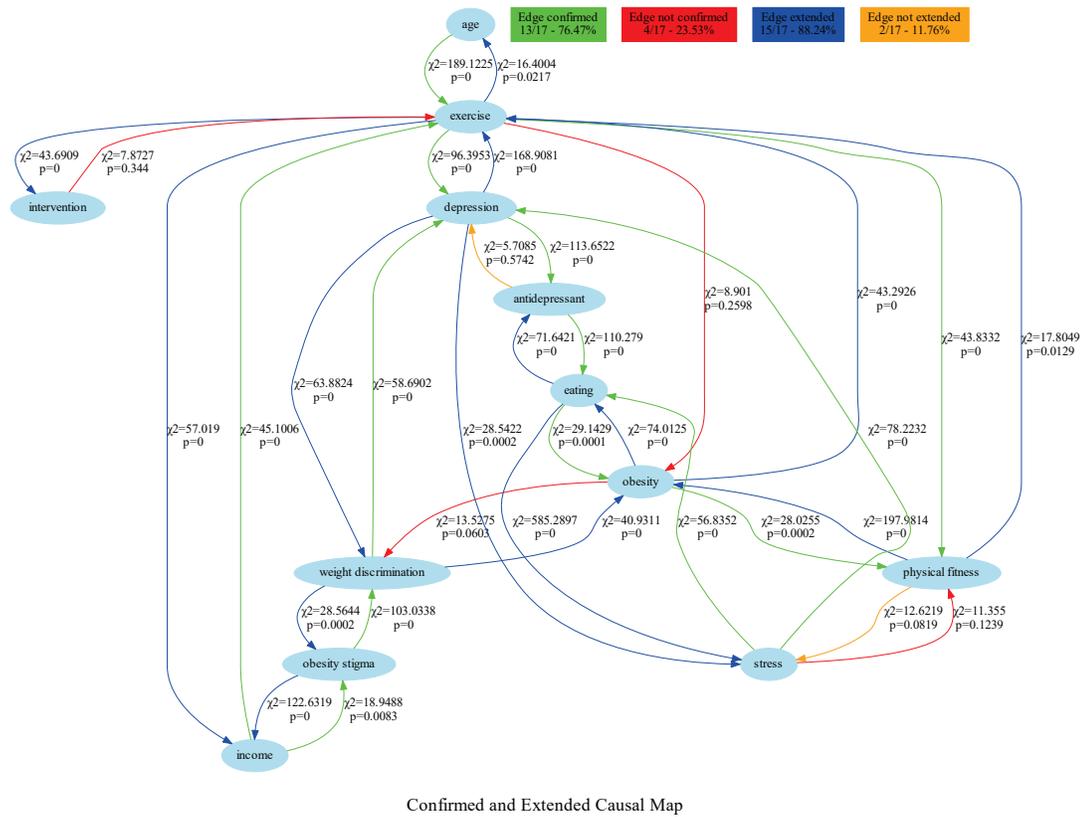


Figure 4.5: Causal Map using Granger causality

In comparison to Figure 4.5 we can observe that the time series for edge (weight discrimination, depression) shown in Figure 4.6 rejects the null hypothesis that weight discrimination does not Granger-causes depression and vice-versa. This proves that mining literature could help gain the strength of evidence for the associations confirmed using our proposed pipeline.

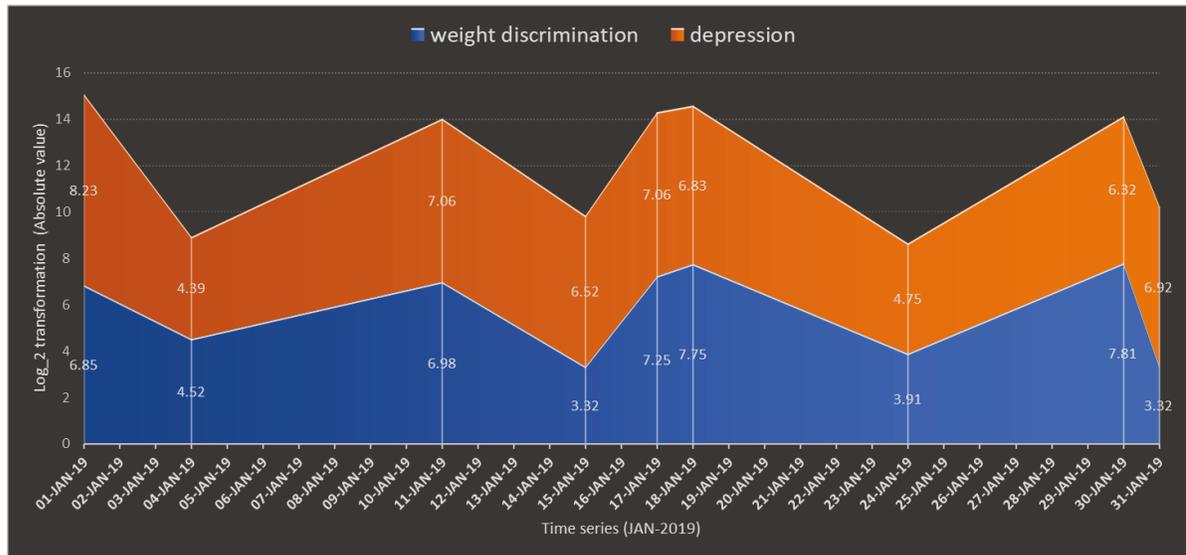


Figure 4.6: Causality plot of edge (weight discrimination, depression) absolute frequency in academic papers for January 2019 month

For Twitter conversations we provide a proof of concept on two edges ('Obesity', 'Stress') and ('Obesity', 'Education'). Since the time frame for this analysis was one week, the maximum lag achieved was 1. Our main focus was on conversations creating constructs, which in the case of edge ('Obesity', 'Stress') [See Figure 4.6] was possible because people were tweeting in regards with stress as the cause of obesity. On the other hand edge, ('Obesity', 'Education') [See Figure 4.7] did not show any constructs because people were not tweeting about education as a cause of obesity. Using constructs from real-time conversations between people is a leap forward to ascertain causality.

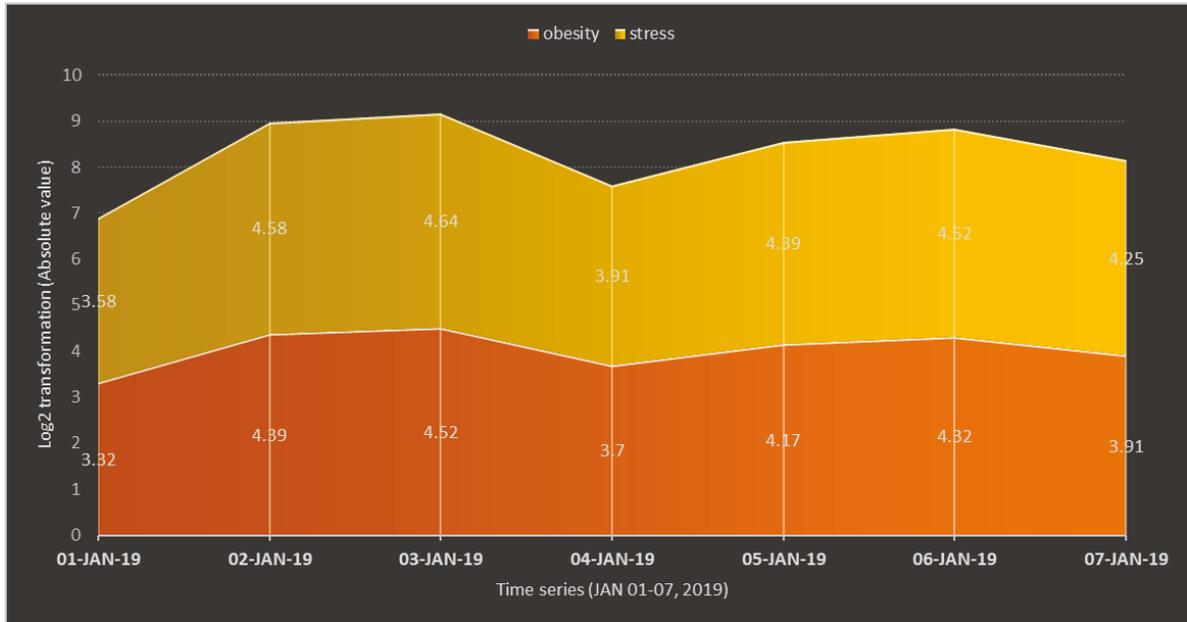


Figure 4.7: Causality plot of edge (obesity, stress) absolute frequency in Twitter conversations for January 01-07 2019 week

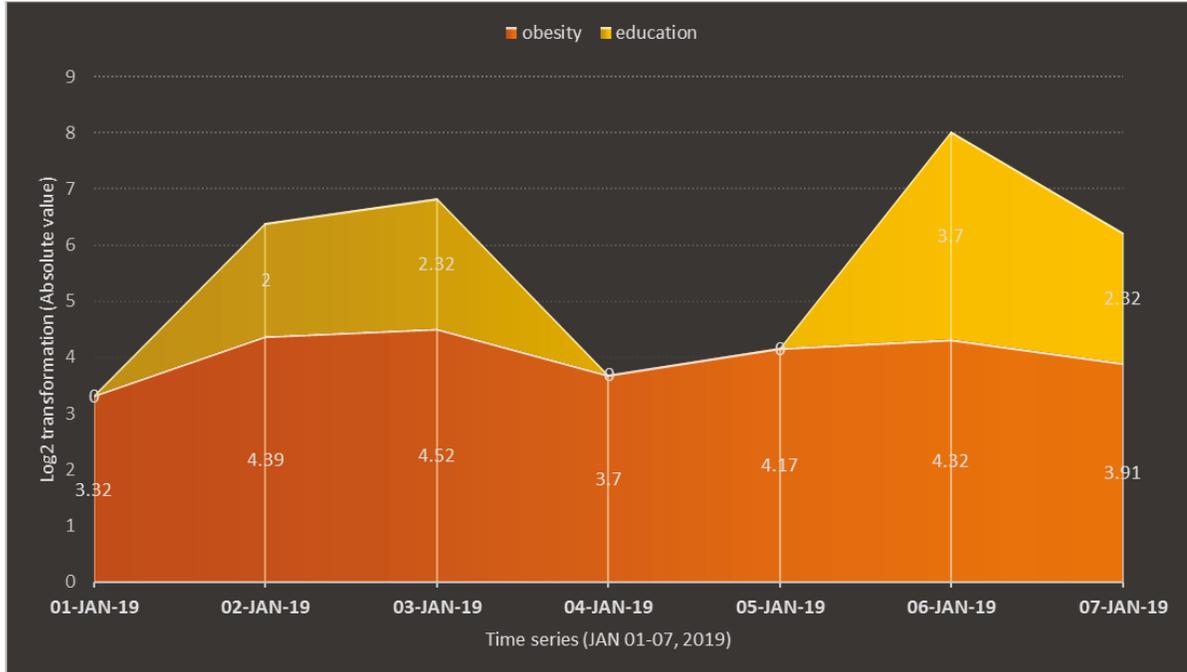


Figure 4.8: Causality plot of edge (obesity, education) absolute frequency in Twitter conversations for January 01-07 2019 week

4.5 Conclusion

4.5.1 Limitations

Using PDF to confirm temporality does not justify the causation results. In comparison to Figure 4.5 where edge (exercise, depression) rejects the null hypothesis that exercise does not Granger-causes depression, and vice-versa, Figure 4.8 time series for that edge (exercise, depression) does not validate the same.

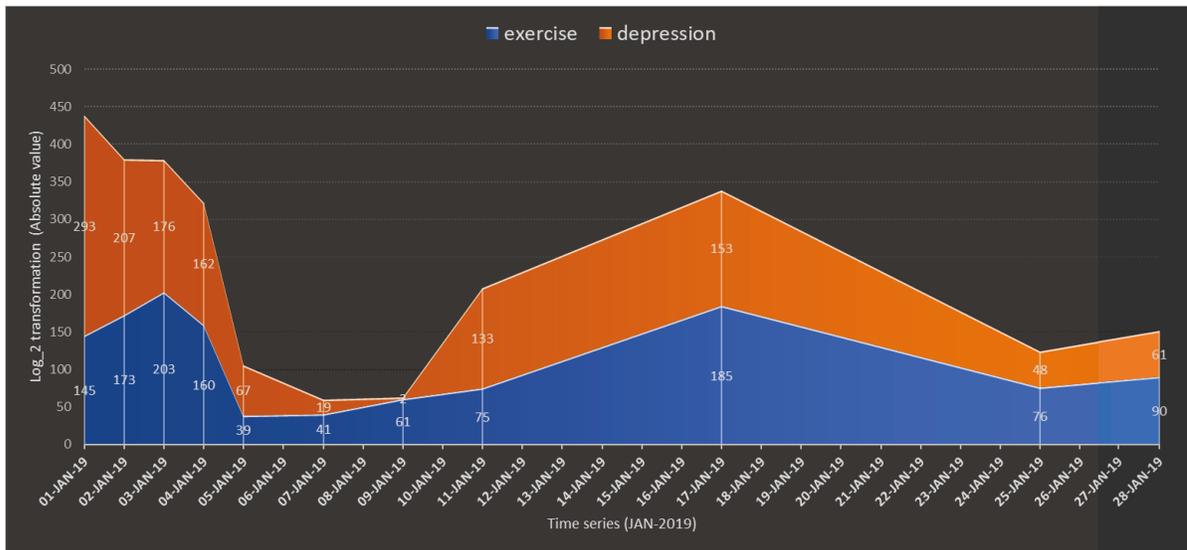


Figure 4.9: Causality plot of edge (exercise, depression) absolute frequency in academic papers for January 2019 month

Also for Figure 4.9 where time series for edge (intervention, exercise) depicts the null hypothesis intervention does not Granger-causes exercise and vice-versa, the Figure 4.5 does not validate the same when we ran `grangercausalitytest` on the given time series.

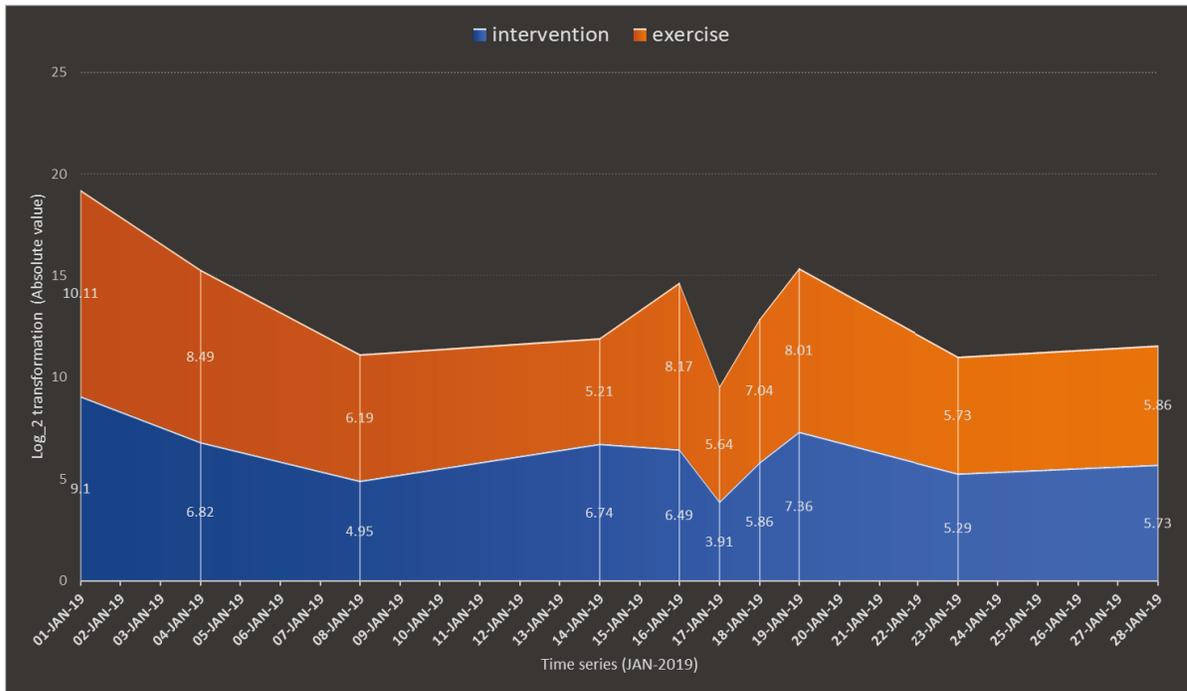


Figure 4.10: Causality plot of edge (intervention, exercise) absolute frequency in academic papers for January 2019 month

4.5.2 Future work

Following temporality of tweets by mining Twitter conversations to create constructs out of it could be a leap forward in ascertaining Granger causality. In this chapter, we have provided a proof of concept that generating time series over the edges of concept map by mining their relevant Twitter conversations can further strengthen the evidence of confirmation and extension of these concept maps. Once we have the causality evidence for our concept map, we further detect the shift in public opinion as well to analyze if these computational results generated using Twitter data are an artifact of our proposed methodologies or accurate reflection of events related to that concept map. The next chapter is a case study on political events which uses our methodology to address the shift in public opinion.

Chapter 5

From Causality to Sarcasm: Mining the Shift of Opinions Regarding the Supreme Court on Twitter

5.1	Abstract	61
5.2	Introduction	62
5.3	Background	64
5.4	Methods	65
	5.4.1 Overview	65
	5.4.2 Data collection	66
	5.4.3 Data cleaning	67
	5.4.4 Analyses	69
5.5	Results	73
5.6	Discussion	76
5.7	Conclusion	77

5.1 Abstract

Twitter is a valuable source for learning about public opinion and political communication, and Twitter mining offers a way to analyze large numbers of tweets to help us understand political associations the public makes. However, the use of incivility and

particularly sarcasm in political discourse may pose a challenge for Twitter mining in the context of politics. In this study, we apply Twitter mining to the 2018 confirmation of Judge Brett Kavanaugh to the Supreme Court to look for possible changes in public opinion of the Court in the wake of the confirmation hearing and to determine whether sarcasm in political messages on Twitter can alter the results of computational methods when using large data-sets. Examining two waves of tweets, one in the days immediately following the confirmation and one a month later, we find evidence of a shift in public opinion as associations between the Supreme Court and partisanship emerge only in the latter period. Using sentiment analysis, we also demonstrate that sarcasm led to over-categorization of positive tweets that altered the results by suggesting the public viewed partisanship on the Supreme Court favorably.

5.2 Introduction

Existing research on Twitter and politics focuses on both elite and public use of Twitter. Elite-centered studies consider which political officials use Twitter, how, and for what purposes. Who uses Twitter has been of particular interest in studies of congressional members [157, 158, 159], and campaign-focused research has looked at gender differences in congressional candidates' Twitter use [160]. The content of elected officials' tweets can be a key to understanding what activities officials prioritize in office (e.g. [161, 162]). Scholars have also relied on tweets to gain insight into the relationship between elected officials and their constituents or followers on Twitter (e.g. [163, 164]). Recognizing journalists' reliance on Twitter, some studies focus on elite Twitter users' influence on media coverage in the context of political campaigns [165] and in governing [166].

Many scholars have turned to Twitter to learn more about the political behavior, opinion, and communication of the public. Three areas of study have gained considerable attention: Twitter in the context of campaigns and in the development of social movements, the effects of Twitter on political interest, and the rhetoric and tone in political tweets. Campaign studies have examined the effectiveness of candidate messages delivered via Twitter compared to messages voters receive from traditional news media [167]. Tweets in campaigns have been analyzed in several ways [168] including to provide evidence of ideological echo chambers [169], compare the spread of fake news and corrections of such stories [170], and study citizens' perceptions of candidates [171] and responses to candidate speeches [172]. Numerous studies exist on the role of Twitter in social movements,

particularly the use of social media to react to events and encourage the formation of collective identities that result in activism (e.g. [173, 174, 175, 176]). And there have been efforts to demonstrate the political effects of Twitter on the public; for example, Bode and Dalrymple [177] found in a survey that Twitter users were more interested in politics and less trusting of mainstream media, suggesting important implications for political communication. Twitter is an excellent source for rhetorical studies with scholars using tweets to examine the incivility, sarcasm, and the use of humor in political discourse [178, 179, 180].

The messages produced on Twitter (i.e., tweets) may be analyzed qualitatively as researchers read them, or they may be automatically examined through a computational lens. Automatically finding patterns in the data produced on social websites is known as *social web mining*, and specifically *Twitter mining* when the social platform of interest is only Twitter [65]. A very large number of tweets is one of the key reasons to perform Twitter mining instead of, or in complement to, a qualitative approach. For instance, the phenomenon of interest may only be observed through a massive amount of data, such as tracking the monthly opinions of various communities in each US state over months. Although Twitter mining may lack the depth or nuance of qualitative methods on individual tweets, coping with data-sets consisting of millions of tweets requires a computational method and special infrastructure [100]. Such large data-sets are either impossible to process for a team of researchers, or would limit us to simple forms of qualitative analyses that can be accurately performed via crowdsourcing. Twitter mining is a well-established approach in health behaviors [26, 27, 71], for instance by analyzing foods consumed using 503 million tweets [72]. Our approach applies Twitter mining to the 2018 confirmation of Judge Brett Kavanaugh. This topic was previously analyzed by Darwish, who used 23 million tweets to analyze the polarization of 687,194 users [181]. While the work of Darwish focuses on polarization around Kavanaugh, we analyze a less direct effect: whether the Kavanaugh case changes public opinion about the Supreme Court itself. Our first research question thus contributes to the literature on politics and media:

(Q1) Is there a shift in the public opinion of the Supreme Court before and after the confirmation of Judge Brett Kavanaugh?

The analysis of political discourse on Twitter has its own challenges, in part due to the use of incivility, sarcasm, and humor. Sarcasm has been the topic of numerous

papers on Twitter mining [182], whose methodologies range from a reliance on self-disclosure of sarcasm (e.g., via the hashtags #sarcasm or #sarcastic [183, 184]) to the use of advanced artificial tools such as neural networks [185]. In this chapter, our second research question is whether sarcasm does matter for Twitter mining in the context of political discourse. Formally, our question is:

- (Q2) Can sarcasm in political messages posted on Twitter significantly alter the results of computational methods even when using large data-sets?

The remainder of this chapter is organized as follows. In section 5.3, we provide a succinct background to the context of our case study, that is, the confirmation of Judge Brett Kavanaugh. Given this context, we introduce our methods in section 5.4, in line with the framework that we recently used in Sandhu *et al.* [103]. In section 5.5, we provide our results on the case study. The final section contextualizes our results with respect to our two research questions, and concludes with suggestions for future work.

5.3 Background

When United States Supreme Court Justice Anthony Kennedy decided to retire in the summer of 2018, the process to confirm Federal Appellate Court Judge Brett Kavanaugh to replace him was inevitably going to be partisan. Democrats were still angry that the Republican majority in the Senate had denied former President Barack Obama the opportunity to appoint a replacement for conservative Justice Antonin Scalia, who had died unexpectedly during Obama’s last year in office. Refusing to hold confirmation hearings or a vote on Obama’s nomination, Republicans delayed until Republican President Trump came to office and nominated a conservative to fill Scalia’s seat, keeping the court closely divided ideologically. Kennedy was the swing vote who often broke the tie between the Court’s four liberals and four conservatives in close cases. Kavanaugh was much more reliably conservative than Kennedy and was expected to move the Court further to the right ideologically, potentially changing the balance on the Court for decades and endangering liberal precedents on abortion and race decided by earlier courts.

The confirmation hearing by the Senate Judiciary Committee took place over four days from September 4 through 7, and the coverage of the hearings focused on the “partisan rancor” among senators on the committee [186]. Democrats complained that the Republican majority had refused to get all of the documents Democrats wanted to see

from Kavanaugh's days of working in the George W. Bush Administration. They raised concerns that Kavanaugh's "expansive view of executive power" would make him an advocate for Trump on the Supreme Court. Republicans accused the Democrats of making the hearings about Trump instead of judicial qualifications [186]. Further dramatizing the partisan stakes were protestors in the audience who periodically disrupted the hearings and were escorted out of the room and arrested. At the end of the contentious hearings, Kavanaugh appeared to be heading to confirmation.

However, on September 12, reports surfaced that a woman had accused Kavanaugh of sexual assault in an incident that allegedly occurred when both were in high school. Christine Blasey Ford had made Democratic Senator Dianne Feinstein aware of her allegation during the summer but had wanted to remain anonymous. As news of the letter appeared in mainstream news media and Ford went public with her account, Republicans accused Democrats of going to any extreme to block Trump's agenda [187]. Following a tense Senate Judiciary Committee hearing on September 27 in which Ford and Kavanaugh were questioned and a short investigation by the FBI that found no conclusive proof of the assault, the Senate voted on October 6 to confirm Kavanaugh on a 50-48 vote with only one Democrat crossing party lines to support the nomination.

The coverage of Kavanaugh's confirmation focused on the partisan process that left members of both parties lamenting the bitterness and division and on the protests and anger among citizens. Those on the left worried about the direction of the Court, and those on the right decried the tactics of the left. That is the context in which the tweets collected for this study were posted. The high profile partisan fighting and the focus on the ideological balance of the Court led us to look at the association between Kavanaugh and Supreme Court and partisan.

5.4 Methods

5.4.1 Overview

To date, research on Twitter and politics has incorporated a variety of methodologies. There have been qualitative analyses of relatively small samples of tweets [159, 165], often combined with other qualitative methods including elite interviews. The majority of quantitative studies rely on content analysis of tweets from varying sample sizes. The

studies often select tweets that include a particular hashtag [174, 176], occur within a specified time period [180], or come from specific political elites. The sample sizes vary from a few hundred to hundreds of thousands. The sample sizes may be limited in part by researchers’ continued reliance on human coding, particularly for more nuanced analyses like the tone, political sentiment, or partisanship of tweets. In contrast, Twitter mining often involves millions of tweets, with examples including the analysis of 23 million tweets by Darwish [181] or several examinations of the 2016 U.S. presidential election either during the campaign announcement stage (4 million tweets [188]) or throughout the election (50 million tweets [189]). In this chapter, we also take a Twitter mining approach to cope with large data-sets. The remainder of this section explains how we collected, cleaned, and analyzed the data.

5.4.2 Data collection

The conceptual model of this study is shown in Figure 5.1, depicting concepts as nodes and association as undirected edges. This model is not merely a diagram or framework: rather, it serves as input to both our data collection and analysis (section 5.5.4). For data collection, the nodes of the model specify which tweets to collect. The justification for each node is as follows. ‘Kavanaugh’ and ‘Supreme Court’ were included as his confirmation to the court is the event of interest. The inclusion of ‘Partisan’ allows us to specifically examine whether the public opinion sees the Supreme Court as a partisan body (Q2). ‘Trump’ nominated Kavanaugh and was often mentioned in connection with the hearings [186], thus he was also included as a key actor. Kavanaugh was a frequent target of the ‘#MeToo’ movement [190, 191, 192], which Trump attacked as a means to support his nominee [193, 194]. Consequently, #MeToo was part of the search terms. Finally, ‘Democrats’ and ‘Republicans’ were included as the two main parties involved in this political event.

Our process consisted of collecting tweets including at least one of (*i*) the seven nodes’ names (Supreme Court, Partisan, Kavanaugh, Trump, Republicans, #MeToo, Democrats), or (*ii*) a variation of the seven nodes’ names. If variations were not included, then we would have an overly restrictive data collection approach, for instance accepting tweets with ‘democrats’ or ‘#MeToo’ but rejecting those containing ‘democrat’ or ‘MeToo’. Manually listing all variations can be prone to omissions, thus introducing a human bias in the data collection process. Consequently, we automatically generate all variations for

each node’s name using the derivationally-related forms from the WordNet library. As a result, we would not only collect tweets if they mentioned ‘partisan’, but also accept ‘partisanship’ or ‘partizan’.

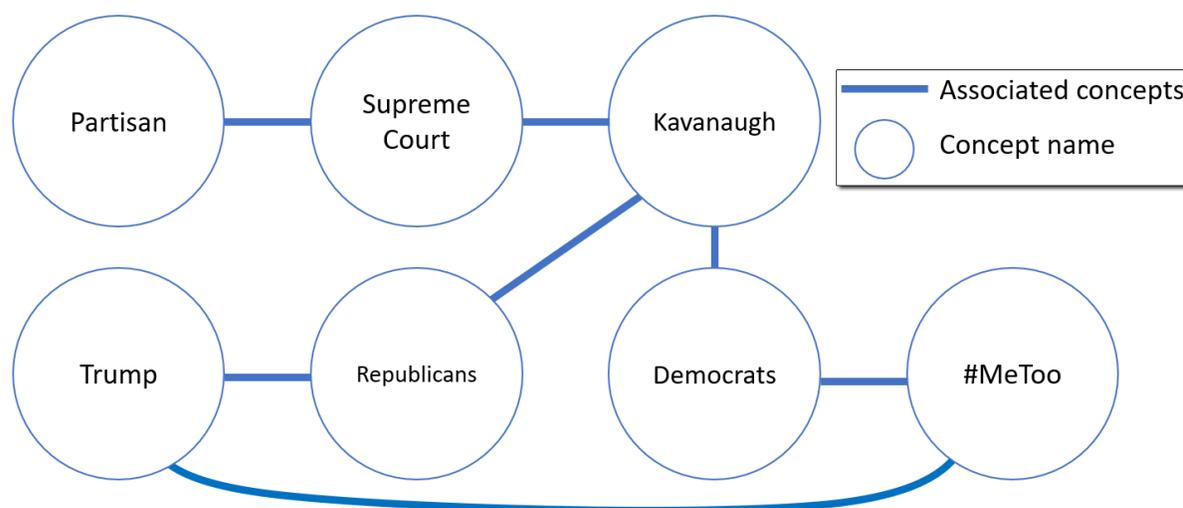


Figure 5.1: Conceptual model for our study. Nodes are used as keywords to retrieve tweets (section 5.5.2) whereas links provide a list of possible associations to examine in the analysis phase (section 5.5.4).

We collected data in two waves. Kavanaugh’s confirmation hearings took place in September, and he was confirmed by the Senate on October 6th, 2018. Our first wave thus covered the period during which there was an intense media coverage of the event, which allows us to observe how the event is influencing the public’s opinion about the Supreme Court. Data collection started on October 9th and finished on October 15th, yielding 14,558,962 raw tweets (i.e. before cleaning). To investigate the longer-term effect of the hearing onto the public’s perceptions of the Supreme Court, the second wave was configured to acquire approximately as many tweets as the first one. It took place from November 11th to 17th and resulted in 14,558,963 raw tweets.

5.4.3 Data cleaning

A large data-set can include content that is not only unnecessary for analysis but may also prove to be detrimental, as noise can detract from accuracy and processing irrelevant data creates a computational burden. Consequently, a large data-set of tweets typically goes through extensive pre-processing before analysis. As the various pre-processing

options and their effects have been described elsewhere [94, 95, 96, 97, 102], we only succinctly describe the ones that we use here. We start with *removing content* that would not be processed in the analysis: URLs, mentions (starting with @), hasthags, numbers, emojis, punctuation, line returns, and extra spaces. We then converted all characters to *lowercase* to prepare for the third step which transforms the content. The *transformation* consisted of expanding acronyms (e.g., ‘?up’ becomes ‘what’s up?’) and contracting words as they are misspelled either involuntarily or for the sake of emphasis (e.g., ‘aaaaand’ turns to ‘and’, ‘fooooood’ becomes ‘food’). Having performed this transformation, we were then able to do a final wave of *content removal* for words that are not considered informative per the English language, such as ‘and’, ‘an’, or ‘the’. The last step performs lemmatization, which consists of simplifying each word to its root form (e.g., ‘eat’ instead of ‘eating’) by doing a morphological analysis of words. This step is an essential component of Natural Language Processing (NLP), as it allows us to map inflected forms of a word into a single base which can be consistently counted (e.g., to establish the frequency of a term in a corpus).

After pre-processing, the data-set from the first wave (October 9-15th) was reduced to 1,696,268 tweets (11.65% of the initial data-set) and the data-set from the second wave (November 11-17th) was reduced to 1,884,486 tweets (12.94% of the initial data-set). Thus reductions in the size of the data-set are a typical outcome of pre-processing. In our recent study on the topic of obesity, we went from 6,633,625 tweets to 1,791,333 [103] (27.00% of the initial data-set). The use of derivationally-related forms (section 5.5.2) is one of the reasons for which many of the tweets were discarded, because it included ‘#’ as a valid alternative for ‘#MeToo’. While we could have manually removed ‘#’ from the data collection, it would have been an ad-hoc alteration of an automatic process. We thus preferred keeping the process fully automatic, knowing that *many* irrelevant tweets would be discarded in a controlled manner during the pre-processing stage. We observed that some irrelevant tweets still appear, but in proportions that cannot have a statistically significant impact on results. For instance, ‘Supreme Court’ was included 43,740 times but only 25 tweets were included for containing the irrelevant word ‘courting’, a single tweet included ‘courtly’, and 479 tweets had ‘court’ without ‘Supreme’.

5.4.4 Analyses

We analyzed the pre-processed data in two ways. First, we examined the presence of the associations listed in Figure 5.1: for instance, is partisanship a theme commonly associated with the Supreme Court? Second, we investigated the sentiments that Twitter users express with regard to these associations. Note that both analyses are applied on our two data-sets (October 9-15th and November 11-17th) such that we can see whether there is a shift in the public opinion (Q1). The presence of sarcasm will also be assessed, in relation with our second research question (Q2).

Our process to examine the presence of associations has been detailed elsewhere [103]. In short, there are two inputs to the process: the conceptual model (Figure 5.1), and the pre-processed set of tweets (section 5.5.3). The process loads the tweets onto the conceptual model. That is, it tells us whether each relation included in the conceptual model is confirmed by the data, and if so, to which extent. The conceptual model thus guides the analysis, for example by specifying that ‘Partisan’ should be explored in relation with the Supreme Court.

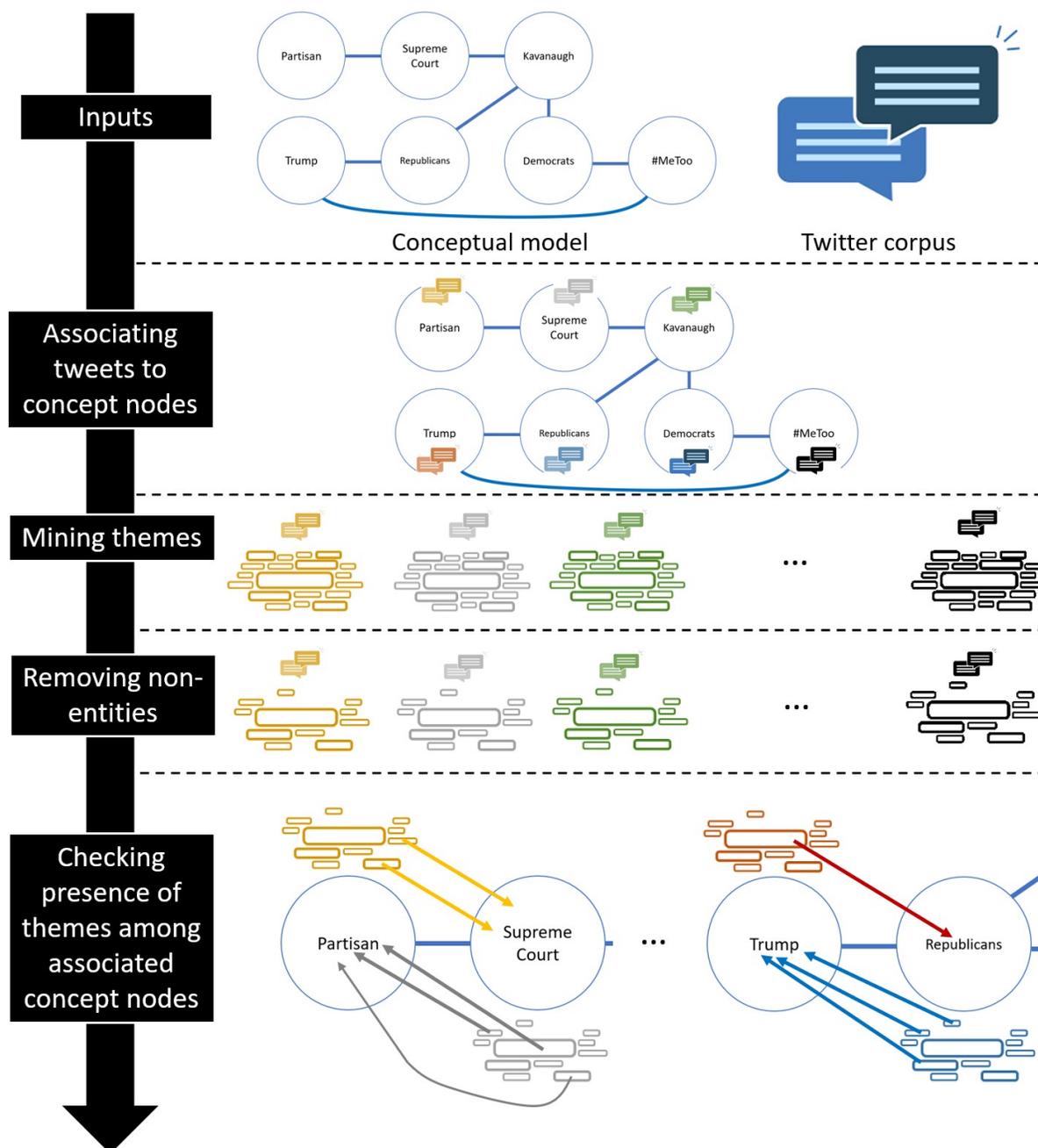


Figure 5.2: High-level view of our analysis process to establish whether, and to which extent, the data-set supports each of the associations in the conceptual model.

A high-level view of the process is provided in Figure 5.2. We start by associating each node with all tweets that contain the node’s name or one of its derivationally related forms. A tweet may be associated with several nodes: for example, the pre-processed tweet *“entire democrat propaganda media machine egregious unethical understatement*

come describe main stream media response national disgrace vile kavanaugh smear stay angry n t complacent punish november” will be associated with both Kavanaugh and Democrats. We continue by extracting the main themes within the set of tweets associated to each node. As every step of our process, this extraction is automatic and relies on Latent Dirichlet Allocation (LDA) [195], which has been used in numerous studies applying Natural Language Processing to Twitter [196, 197]. Since our conceptual model articulates relations of interest between *entities*, we filter the themes with Google Cloud Natural Language APIs to focus on the entities with which each node may be associated. If the associated themes for a given node correspond to connected nodes in the conceptual model, then the corpus supports to this connection. For instance, if ‘partisanship’ is a theme for the corpus relevant to ‘Supreme Court’, then it contributes to the evidence-base for a connection between ‘Supreme Court’ and ‘Partisan’.

Results from the process aforementioned do not only depend on the two inputs (conceptual model and pre-processed tweets) but also on three parameters, whose values must collectively be set such that results are optimal. This is a process known as *hyperparameter optimization*. We optimize using the common *grid search* approach consisting of (i) defining a set of possible values for each parameter, and (ii) computing the results for all combinations of parameter values. The three parameters are related to extraction of themes using Latent Dirichlet Allocation: they are the number of themes (n_{themes}), the number of words per theme (n_{words}), and the frequency to retain significant results ($freq$). The value of $freq$ is limited by the data, as it goes up to the most frequent item. In both data-sets, we varied $freq$ from 2 to 9 by increments of 1. The values of n_{themes} and n_{words} are set similarly to previous work [103], ranging from 5 to 50 by increments of 5. The output from the grid search is shown on Figure 5.3, and we selected the parameter values that maximized the number of associations confirmed.



Figure 5.3: We optimized the values of our three parameters by maximizing the number of associations confirmed on the October data-set (4 out of 7).

Our analysis of sentiments was performed at two levels: we analyzed the sentiments regarding key entities (i.e. the nodes of our conceptual model in Figure 5.1), and we analyzed the sentiments when such entities were associated (i.e. the edges of our conceptual model). Analyzing entities is the most straightforward part: for each entity, we find all relevant tweets through the same first step as in Figure 5.2, and we use the sentiment analysis tool **VADER** (Valence Aware Dictionary and sEntiment Reasoner) to count the percentage of tweets that are positive, neutral, or negative. As an association involves two entities, we get the *union* of tweets relevant to the two entities (which eliminates duplicates) and then we use **VADER** to categorize them. For the sake of efficiency, we do *not* retrieve tweets for one entity, tweets for the other, and spend unnecessary computations to identify and remove replicated tweets. Rather than eliminating duplicates from two independent searches, we perform a single search for which the keywords are gathered from both entities using derivationally related forms.

5.5 Results

Our two analyses examined the presence of associations and then the types of sentiments (section 5.5.4). The results for both analyses are presented in turn. We confirmed 4 out of 7 potential associations (57%) on the October data-set, and 5 out of 7 (71%) on the November data-set (Figure 5.4). Most importantly, the association between the Supreme Court and Partisan was confirmed in November but not in October. That is, after the confirmation of Judge Brett Kavanaugh, the Supreme Court was associated with the idea of partisanship. This addresses our first research question, thus presenting evidence for a shift in the public opinion of the Supreme Court. We also note that the association of Judge Brett Kavanaugh with the Democrats in October has stopped in November.

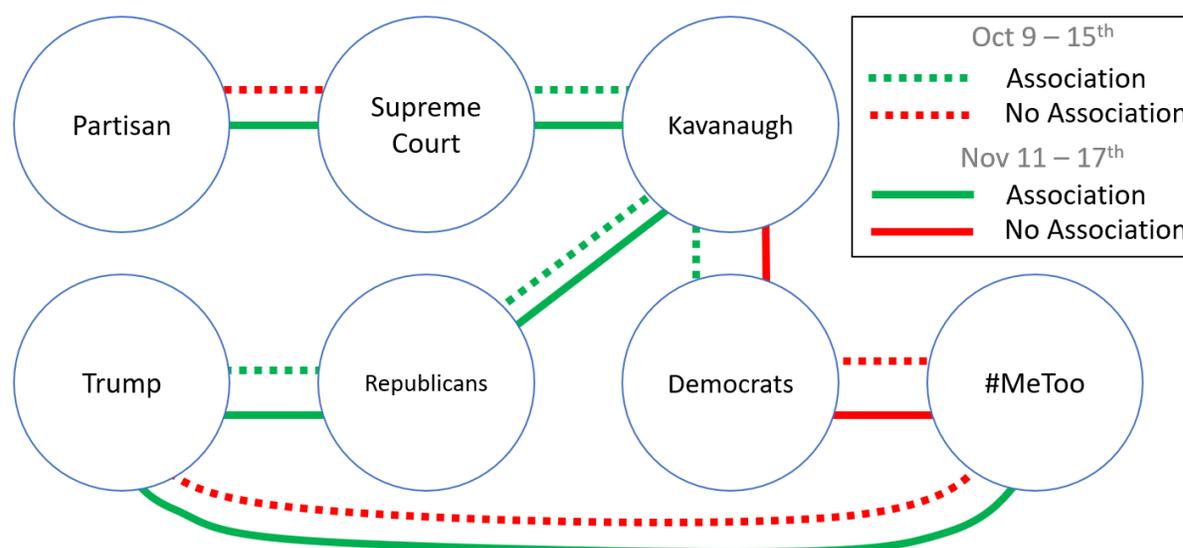


Figure 5.4: Associations that were confirmed (green) or not (red) for both of the data-sets. Note that the association between the Supreme Court and Partisan is confirmed in November but not in October.

The other research question examines whether sarcasm in the tweets can alter the results. We note that associations cannot be altered by sarcasm, because they are only about the joint *presence* of entities rather than the emotions expressed regarding these entities. The primary impact of sarcasm is thus measured through our analysis of sentiments. From Figure 5.5, we observe that the Supreme Court and partisanship evoke a clearly positive feeling. As it would seem unlikely for a population to endorse prejudice in the highest federal judiciary court, we explored the roots of this result. A manual inspection from several pre-processed tweets categorized as positive reveals the likely use of sarcasm.

For instance, several tweets may not truly be thankful (e.g., “thanks strictly partisan supreme court vote”, “Thanks republican [for a] partisan supreme court [!]”) or happy (e.g., “Yeah happy put partisan rapist supreme court opinion worth nothing”). The use of words such as ‘thanks’, ‘happy’, ‘great’ (“great non partisan supreme court huh”) or ‘highly’ (e.g., “gop rammed highly partisan misogynistic fratboy supreme court sh”) may thus derail the sentiment analysis. Since our observations were made on a small sample, it is necessary to assess more comprehensively whether sarcasm is prevalent throughout our data-set consisting of approximately 3 million tweets in total. Several solutions have been presented at the Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. In our work, we use the neural network approach of Ghosh and Veale presented at the 7th edition of the workshop in 2016. Their solution allows us to automatically and accurately detect the presence of sarcasm, as it has shown an accuracy of almost 92% on tweets [198]. Results on Figure 5.6 confirm both (i) the presence of sarcasm and (ii) the role it plays in the over-categorization of positive tweets. With regard to the first point, we observe that associations between entities have 36% to 41% of sarcastic posts. The latter is demonstrated as sarcastic posts are *always* more prevalent with positive than negative tweets. We observed the largest difference on tweets involved in the association between Supreme Court and Partisan, where 19% of ‘positive’ tweets were sarcastic versus only 4% for ‘negative’ tweets.

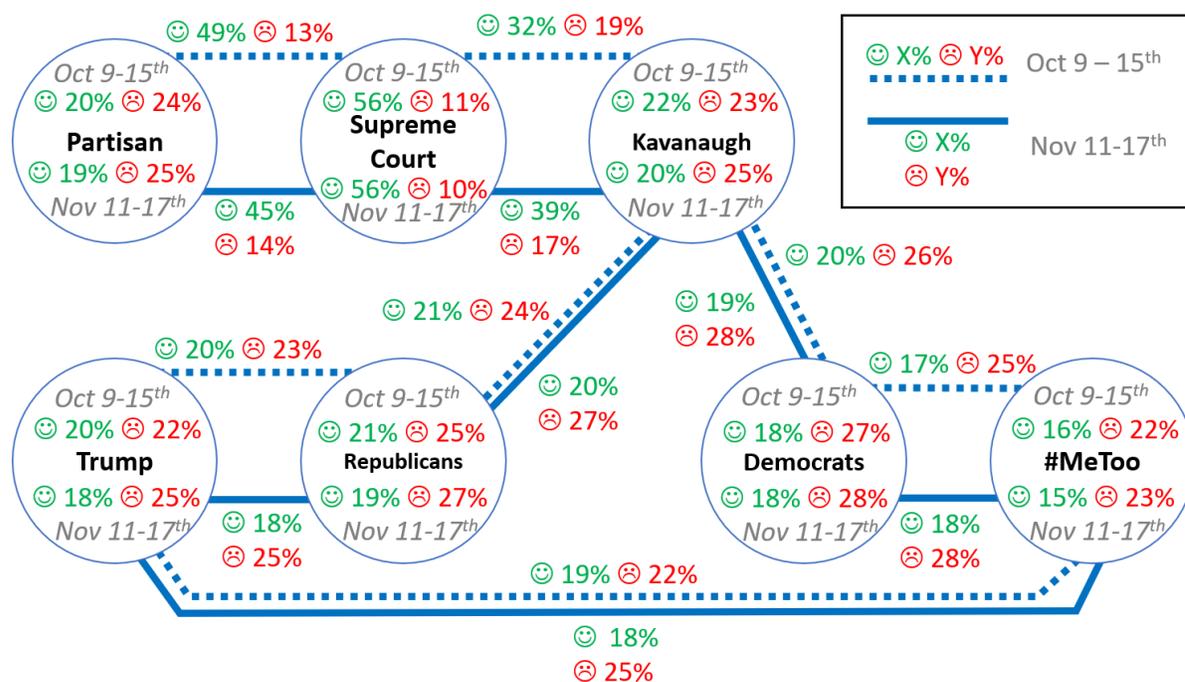


Figure 5.5: Sentiment analysis at both the node- and edge-level. Only the positive and negative categories are shown; all other posts are categorized as neutral.

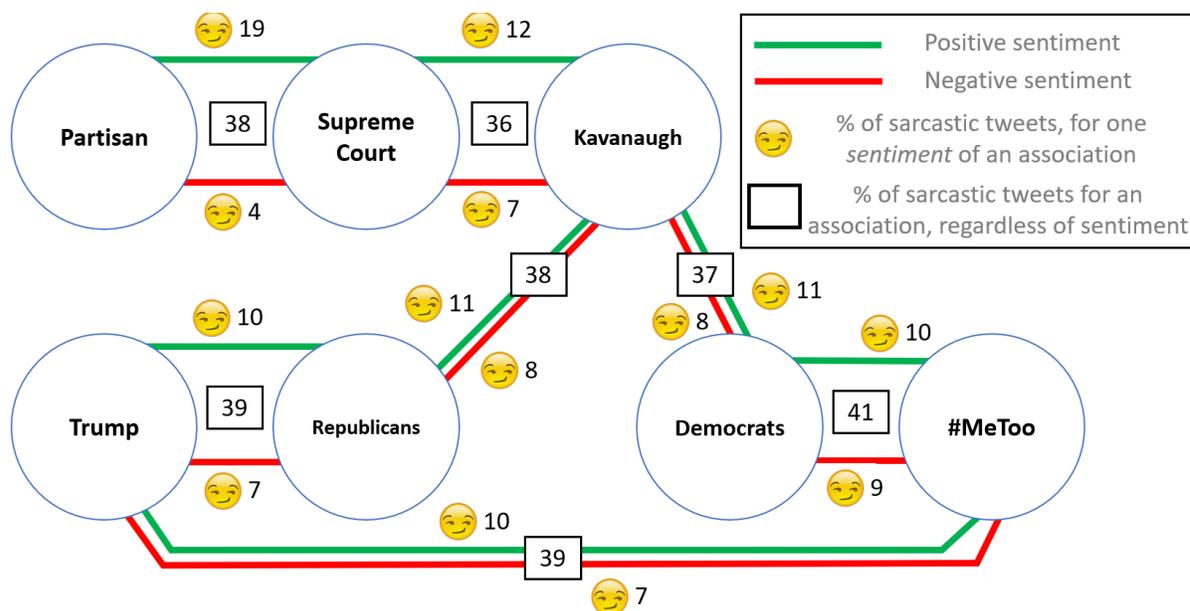


Figure 5.6: Prevalence of sarcasm at the association-level and at the sentiment-level for the Oct 9-15th data-set. Sarcasm for tweets with neutral sentiments is not displayed. Sarcasm is only computed on tweets with at least four words using the solution of Ghosh and Veale [198].

5.6 Discussion

Twitter has been analyzed in numerous studies to further our understanding of the political behavior, opinion, and communication of the public. While qualitative analyses have demonstrated their usefulness in capturing the nuance of tweets, automatic approaches (i.e., ‘Twitter mining’) can provide a useful complement to analyze large data-sets formed of millions of tweets. A challenge in this situation is the use of incivility, sarcasm, and humor. Our work provides a technical contribution to document the impact of sarcasm on large scale analyses. We confirm that sarcasm did alter the results of computational methods, as it led to the misleading suggestion that the public likes bias in the Supreme Court. Our study also makes a contribution to the literature on politics and media through our specific case study of tweets regarding the confirmation of Judge Brett Kavanaugh to the Supreme Court. An association between the Supreme Court and Partisanship only after, and not during, the confirmation. This suggests that the confirmation has modified the views that the public holds regarding the Supreme Court, and hence more broadly with the partiality of justice in the United States.

Our study has several strengths. First, we rely on methods and implementations that have been previously published. **VADER** has been used for sentiment analysis of a Twitter corpus in several studies published within the last three years [199, 200, 201]. Similarly, our approach to finding associations uses the Latent Dirichlet Allocation (LDA), a well-established method to mine themes from text [195], whose accuracy makes it a common choice for applications to Twitter data [196, 197]. As we recently designed the process to find associations, it has only been used in our other 2019 study [103]. However, we took a very conservative approach by selecting parameter values that maximize the number of associations.

Although maximizing the number of associations provides a very conservative estimate, it is worth asking whether the lack of an association (e.g., between Supreme Court and Partisan in October 9-15th) is an artifact of our method or an accurate reflection of events. For instance, our results did not find associations between Democrats and the #MeToo movement although other studies have presented tweets that simultaneously refer to both of these entities (c.f., [202], p. 14). As any analysis of a sample, our findings cannot demonstrate the absolute absence of evidence for an association. Rather, our analysis concludes that, in two samples of over 1.5 million tweets, such associations

did not have enough statistical significance to either (i) be included in a theme by Latent Dirichlet Allocation (LDA), or (ii) achieve a sufficient frequency to be retained among significant results. As aforementioned, LDA is a strength of this study and would not be a reason that a theme was missing. Our criteria for ‘sufficient frequency’ are unlikely to provide such a reason either. Indeed, evaluating a ‘sufficient frequency’ used the term-frequency inverse document-frequency (tf-idf), which is one of the most popular information retrieval metrics of the importance of a word *within a corpus*. A word does not need to appear frequently in general (e.g., ‘thus’, ‘then’) because the *inverse document-frequency* increases the weight of terms that are frequent within some documents rather than throughout a collection. In an extreme case, the tf-idf measure may be a reason to have associations that do *not* exist (false positive) rather than missing existing ones (false negative), as the measure has been noted for being occasionally lenient by accepting irrelevant terms if a collection was noisy [203]. In sum, the design of our study gives credit to the idea that the Supreme Court was only associated with partisanship on November 11-17th but not on October 9-15th.

5.7 Conclusion

By automatically analyzing millions of tweets, we find evidence of a shift in the public opinion. Such as in this case study of the Supreme Court before and after the confirmation of Judge Brett Kavanaugh. The Supreme Court was associated with partisanship one month after the confirmation but not in the days that immediately followed. We also demonstrate that sarcasm in political tweets can significantly alter the outcome of tweet mining even when using large data-sets.

Chapter 6

Discussion, Future Work & Conclusion

6.1	Overview	78
6.2	Scope For Improvement & Current Exceptions	78
6.3	Conclusion	79
6.4	Availability of data and material	80

6.1 Overview

Both social media data and expert reports may be used to take into account popular perspectives and expert opinions when creating large conceptual models. The pipeline to validate complex conceptual models by mining Twitter data can be used as a generic methodology for validating conceptual models using citizen science approach. Once a model is validated , it can be further extended using Granger causality tests. By automatically analyzing millions of tweets, we find evidence of a shift in the public opinion.

6.2 Scope For Improvement & Current Exceptions

For improvement, we changed from old LDA model to LDA multicore model and then to LDA multicore GPU model(using tensorflow). It took 10 days using old LDA model,

5 days using LDA multicore model, 4 hours using LDA multicore GPU model (using tensorflow) for one round of our parameters number of topics, number of words per topic and maximum tf-idf weight using Twitter data on our concept map (see Figure 3.2). Also while using LDA multicore GPU model (using tensorflow) for one round on millions of tweets but small maps like the political map (7 nodes(refer Chapter 4)) it took 2 hours of computation only. LDA is as fast as it can be, but we are using it to run multiple times (depending on number of nodes of usermap) which is what takes time. And then each node contains millions of tweets which adds on to more processing time. There is scope for improvement in future by considering parallel LDA processing for each node to fetch these themes much more faster in future. Our LDA model is very much efficient with the constraints(refer table B.1) and systems(GPU) we used on. Getting results in 4 hours after running LDA for 177(nodes(refer Chapter 2)) times on one round of n_topics, n_words, max_df is a milestone we overcame. It will take much less time if we find themes for single node.

6.3 Conclusion

In this thesis, an approach to validate conceptual models is presented. This approach is applied to extract relevant themes for each concept by mining Twitter data which is being used for further analysis to support the validation of connection between concepts of our concept map using our algorithms. To start with the analysis, we input Twitter data-set fetched using the concept names of the map and its derivationally related form and user concept map in JSON and CSV format which are converted to dictionaries when used in our algorithms..

The whole process is broken down into multiple steps, the two main being detecting themes and mapping it with related concept by comparing with derivationally related form of that concept. The purpose of extracting themes is to discover the abstract topics that occur in a collection of documents here set of relevant tweets corresponding to concept name. The mapping it back to relevant concept name is finally done to validate the conceptual model.

6.4 Availability of data and material

The data-sets and conceptual model generated during and/or analysed during the current study are available from the thesis author.

Appendix A

List of Abbreviations

- **API** Application Program Interface
- **NLTK** Natural Language ToolKit
- **CPU** Central Processing Unit
- **GPU** Graphic Processing Unit
- **PHSA** Provonvincial Health Services Authority
- **PDF** Portable Document Format
- **LDA** Latent Dirichlet allocation
- **CSV** Comma-separated values
- **NLP** Natural language processing
- **ASCII** American Standard Code for Information Interchange
- **SD** System Dynamics
- **FCM** Fuzzy Cognitive Maps

Appendix B

Values For LDA Multicore model Parameters

Table B.1: LDA model parameters

Parameters	Value
corpus	corpora.Dictionary.doc2bow(word)
id2word	corpora.Dictionary(word)
num_topics	n_topics(user-defined global parameter)
n_words	user-defined global parameter for number of words per topic
chunksize	10000
passes	20
eval_every	None
workers	3
dtype	np.float64
iterations	400

Bibliography

- [1] Rick Bonney, Heidi Ballard, et al. Public participation in scientific research: Defining the field and assessing its potential for informal science education. a cause inquiry group report. *Online Submission*, 2009.
- [2] Peter O Dunn and David W Winkler. Climate change has affected the breeding date of tree swallows throughout north america. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1437):2487–2490, 1999.
- [3] Christine Overdevest and Brian Mayer. Harnessing the power of information through community monitoring: insights from social science. *Tex. L. Rev.*, 86: 1493, 2007.
- [4] Jennifer Shirk, Heidi Ballard, et al. Public participation in scientific research: a framework for deliberate design. *Ecology and society*, 17(2), 2012.
- [5] Suzanne B Cashman, Sarah Adeky, et al. The power and the promise: working with communities to analyze data, interpret findings, and get to outcomes. *American journal of public health*, 98(8):1407–1417, 2008.
- [6] Michael Paolisso and Jeremy Trombley. Cognitive, material and technological considerations in participatory environmental modeling. In *Environmental modeling with stakeholders*, pages 3–23. Springer, 2017.
- [7] Martin Davies. Concept mapping, mind mapping and argument mapping: what are the differences and do they matter? *Higher education*, 62(3):279–301, 2011.
- [8] Chris Goodier, Simon Austin, et al. Causal mapping and scenario building with

- multiple organisations. *Futures*, 42(3):219–229, 2010.
- [9] Philippe J Giabbanelli and Andrew A Tawfik. Overcoming the pbl assessment challenge: Design and development of the incremental thesaurus for assessing causal maps (itacm). *Technology, Knowledge and Learning*, pages 1–8, 2018.
- [10] Joseph D Novak. *Learning, creating, and using knowledge: Concept maps as facilitative tools in schools and corporations*. Routledge, 2010.
- [11] Rebecca Jordan, Alycia Crall, et al. Citizen science as a distinct field of inquiry. *Bioscience*, 65(2):208–211, 2015.
- [12] Steven Gray, Rebecca Jordan, et al. Combining participatory modelling and citizen science to support volunteer conservation action. *Biological conservation*, 208:76–86, 2017.
- [13] Beykikhoshk A, Arandjelovi O, et al. Using twitter to learn about the autism community. *Social Network Analysis and Mining*, 22(5(1)), 2015 Dec 1.
- [14] Ji X, Chun SA, et al. Twitter sentiment classification for measuring public health concerns. *Social Network Analysis and Mining*, 13(5(1)), 2015 Dec 1.
- [15] Fink C, Schmidt A, et al. Complex contagions and the diffusion of popular twitter hashtags in nigeria. *Social Network Analysis and Mining*, 1(6(1)), 2016 Dec 1.
- [16] Philippe J Giabbanelli, Thomas Torsney-Weir, and Vijay Kumar Mago. A fuzzy cognitive map of the psychosocial determinants of obesity. *Applied soft computing*, 12(12):3711–3724, 2012.
- [17] Lydia Drasic and Philippe J Giabbanelli. Exploring the interactions between physical well-being, and obesity. *Canadian Journal of Diabetes*, 39:S12–S13, 2015.
- [18] Brynle Owen et al. Understanding a successful obesity prevention initiative in children under 5 from a systems perspective. *PloS one*, 13(3):e0195141, 2018.

- [19] Jaimie McGlashan et al. Quantifying a systems map: network analysis of a childhood obesity causal loop diagram. *PloS one*, 11(10):e0165459, 2016.
- [20] Jaimie McGlashan et al. Comparing complex perspectives on obesity drivers: action-driven communities and evidence-oriented experts. *Obesity science & practice*, 2018.
- [21] Steven Allender et al. A community based systems diagram of obesity causes. *PLoS One*, 10(7):e0129683, 2015.
- [22] Brian G Giles et al. Integrating conventional science and aboriginal perspectives on diabetes using fuzzy cognitive maps. *Social science & medicine*, 64(3):562–576, 2007.
- [23] Alexey Voinov et al. Tools and methods in participatory modeling: Selecting the right tool for the job. *Environmental Modelling & Software*, 109:232–255, 2018.
- [24] Thrishma Reddy, Philippe J. Giabbanelli, and Vijay K. Mago. The artificial facilitator: guiding participants in developing causal maps using voice-activated technologies. In *International Conference on Augmented Cognition*, 2019.
- [25] Jiyeon So et al. What do people like to “share” about obesity? a content analysis of frequent retweets about obesity on twitter. *Health Communication*, 31(2):193–206, 2016.
- [26] Wen-ying Sylvia Chou, Abby Prestin, and Stephen Kunath. Obesity in social media: a mixed methods analysis. *Translational behavioral medicine*, 4(3):314–323, 2014.
- [27] George Shaw Jr and Amir Karami. Computational content analysis of negative tweets for obesity, diet, diabetes, and exercise. *Proceedings of the Association for Information Science and Technology*, 54(1):357–365, 2017.
- [28] Amir Karami et al. Characterizing diabetes, diet, exercise, and obesity comments on twitter. *International Journal of Information Management*, 38(1):1–6, 2018.

- [29] Philippe J Giabbanelli, Jean Adams, and Venkata Sai Pillutla. Feasibility and framing of interventions based on public support: leveraging text analytics for policymakers. In *International Conference on Social Computing and Social Media*, pages 188–200. Springer, 2016.
- [30] Centers for Disease Control and Prevention (CDC). Selected health conditions and risk factors, by age: United states, selected years 1988-1994 through 2015-2016. URL <https://www.cdc.gov/nchs/hus/contents2017.htm#053>.
- [31] Miguel Peralta et al. Prevalence and trends of overweight and obesity in older adults from 10 european countries from 2005 to 2013. *Scandinavian journal of public health*, page 1403494818764810, 2018.
- [32] Jeanne Lubbe. Obesity and metabolic surgery in south africa. *South African Gastroenterology Review*, 16(1):23–28, 2018.
- [33] Youfa Wang, L Wang, and W Qu. New national data show alarming increase in obesity and noncommunicable chronic diseases in china. *European journal of clinical nutrition*, 71(1):149, 2017.
- [34] Marie Ng et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the global burden of disease study 2013. *The lancet*, 384(9945):766–781, 2014.
- [35] Asheley Cockrell Skinner et al. Prevalence of obesity and severe obesity in us children, 1999–2016. *Pediatrics*, page e20173459, 2018.
- [36] Sara N Bleich et al. Interventions to prevent global childhood overweight and obesity: a systematic review. *The Lancet Diabetes & Endocrinology*, 6(4):332–346, 2018.
- [37] Hutchesson MJ et al. ehealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. *Obesity reviews*, 16(5):376–392, 2015.

- [38] Tamim Rajjo et al. Treatment of pediatric obesity: an umbrella systematic review. *The Journal of Clinical Endocrinology & Metabolism*, 102(3):763–775, 2017.
- [39] Pedro J Teixeira et al. Successful behavior change in obesity interventions in adults: a systematic review of self-regulation mediators. *BMC medicine*, 13(1):84, 2015.
- [40] National Institute for Health and Care Excellence. Managing overweight and obesity in adults—lifestyle weight management services. *NICE Public Health Guideline 53*, 2014.
- [41] George Blackburn. Effect of degree of weight loss on health benefits. *Obesity research*, 3(S2):211s–216s, 1995.
- [42] David S Fink and Katherine M Keyes. Wrong answers: when simple interpretations create complex problems. *Systems Science and Population Health*, pages 25–36, 2017.
- [43] Sarah Frood et al. Obesity, complexity, and the role of the health system. *Current obesity reports*, 2(4):320–326, 2013.
- [44] Diane T Finegood. The complex systems science of obesity. In *The Oxford handbook of the social science of obesity*. 2011.
- [45] Harry Rutter et al. The need for a complex systems model of evidence for public health. *The Lancet*, 390(10112):2602–2604, 2017.
- [46] Philippe J Giabbanelli. Analyzing the complexity of behavioural factors influencing weight in adults. In *Advanced Data Analytics in Health*, pages 163–181. Springer, 2018.
- [47] Penny Deck, Philippe Giabbanelli, and Diane T Finegood. Exploring the heterogeneity of factors associated with weight management in young adults. *Canadian journal of diabetes*, 37:S269–S270, 2013.

- [48] Jebb SA, Kopelman P, and Butland B. Executive summary: foresight tackling obesity: future choices project. *Obesity Reviews*, 8:vi–ix, 2007.
- [49] Xue H et al. Applications of systems modelling in obesity research. *Obesity Reviews*, 19(9):1293–1308, 2018.
- [50] Leah Frerichs et al. Mind maps and network analysis to evaluate conceptualization of complex issues: A case example evaluating systems science workshops for childhood obesity prevention. *Evaluation and program planning*, 68:135–147, 2018.
- [51] Lee M Johnston, Carrie L Matteson, and Diane T Finegood. Systems science and obesity policy: a novel framework for analyzing and rethinking population-level planning. *American journal of public health*, 104(7):1270–1278, 2014.
- [52] Laurette Dubé et al. Convergent innovation in food through big data and artificial intelligence for societal-scale inclusive growth. *Technology Innovation Management Review*, 8(2), 2018.
- [53] Srivardhini K Jha, Richard Gold, and Laurette Dube. Convergent innovation platform to address complex social problems: A tiered governance model. In *Academy of Management Proceedings*, volume 2016, page 15150. Academy of Management Briarcliff Manor, NY 10510, 2016.
- [54] Diane T Finegood, Thomas DN Merth, and Harry Rutter. Implications of the foresight obesity system map for solutions to childhood obesity. *Obesity*, 18(S1): S13–S16, 2010.
- [55] Philippe J Giabbanelli and Magda Baniukiewicz. Navigating complex systems for policymaking using simple software tools. In *Advanced Data Analytics in Health*, pages 21–40. Springer, 2018.
- [56] Philippe J. Giabbanelli et al. developing technology to support policymakers in taking a systems science approach to obesity and well-being. *Obesity Reviews*, 17: 194–195, 2016.

- [57] Jenine K Harris et al. Communication about childhood obesity on twitter. *American journal of public health*, 104(7):e62–e69, 2014.
- [58] Janet A Lydecker et al. Does this tweet make me look fat? a content analysis of weight stigma on twitter. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, 21(2):229–235, 2016.
- [59] Joy L Lee et al. What are health-related users tweeting? a qualitative content analysis of health-related users and their messages on twitter. *Journal of medical Internet research*, 16(10), 2014.
- [60] Khalid A Alnemer et al. Are health-related tweets evidence based? review and analysis of health-related tweets on twitter. *Journal of medical Internet research*, 17(10), 2015.
- [61] Jennie C De Gagne et al. Uncovering cyberincivility among nurses and nursing students on twitter: A data mining study. *International journal of nursing studies*, 89:24–31, 2019.
- [62] Christine N May et al. Weight loss support seeking on twitter: the impact of weight on follow back rates and interactions. *Translational behavioral medicine*, 7(1):84–91, 2016.
- [63] Gabrielle M Turner-McGrievy and Michael W Beets. Tweet for health: using an online social network to examine temporal trends in weight loss-related posts. *Translational behavioral medicine*, 5(2):160–166, 2015.
- [64] Zhixian Sui et al. Recent trends in intensive treatments of obesity: is academic research matching public interest? *Surgery for Obesity and Related Diseases*, 2019.
- [65] Daniel E O’Leary. Twitter mining for discovery, prediction and causality: Applications and methodologies. *Intelligent Systems in Accounting, Finance and Management*, 22(3):227–247, 2015.
- [66] Maged N Kamel Boulos et al. Social web mining and exploitation for serious

- applications: Technosocial predictive analytics and related technologies for public health, environmental and national security surveillance. *Computer methods and programs in biomedicine*, 100(1):16–23, 2010.
- [67] Michael J Paul and Mark Dredze. You are what you tweet: Analyzing twitter for public health. *Icwsn*, 20:265–272, 2011.
- [68] Johannes C Eichstaedt et al. Psychological language on twitter predicts county-level heart disease mortality. *Psychological science*, 26(2):159–169, 2015.
- [69] David Ediger et al. Massive social network analysis: Mining twitter for social good. In *2010 39th International Conference on Parallel Processing*, pages 583–593. IEEE, 2010.
- [70] Marika Tiggemann et al. Tweeting weight loss: A comparison of# thinspiration and# fitspiration communities on twitter. *Body image*, 25:133–138, 2018.
- [71] Aron Culotta. Estimating county health statistics with twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1335–1344. ACM, 2014.
- [72] Sofiane Abbar, Yelena Mejova, and Ingmar Weber. You tweet what you eat: Studying food consumption through twitter. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3197–3206. ACM, 2015.
- [73] Sharon E Alajajian et al. The lexicocalorimeter: Gauging public health through caloric input and output on social media. *PloS one*, 12(2):e0168893, 2017.
- [74] Quynh C Nguyen et al. Building a national neighborhood dataset from geotagged twitter data for indicators of happiness, diet, and physical activity. *JMIR public health and surveillance*, 2(2), 2016.
- [75] Paul I Eke. Using social media for research and public health surveillance. *Journal of dental research*, 90(9):1045, 2011.

- [76] Rajesh Patel et al. Social media use in chronic disease: a systematic review and novel taxonomy. *The American journal of medicine*, 128(12):1335–1350, 2015.
- [77] Lauren E Charles-Smith et al. Using social media for actionable disease surveillance and outbreak management: a systematic literature review. *PloS one*, 10(10): e0139701, 2015.
- [78] Molly E Waring et al. Social media and obesity in adults: a review of recent research and future directions. *Current diabetes reports*, 18(6):34, 2018.
- [79] Alexandra Penn. Moving from overwhelming to actionable complexity in population health policy: Can alife help?, 2018.
- [80] Eric Silverman. Bringing alife and complex systems science to population health research. *Artificial life*, 24(3):220–223, 2018.
- [81] Philippe J Giabbanelli and Rik Crutzen. Using agent-based models to develop public policy about food behaviours: Future directions and recommendations. *Computational and mathematical methods in medicine*, 2017, 2017.
- [82] Philippe Giabbanelli and Rik Crutzen. An agent-based social network model of binge drinking among dutch adults. *Journal of Artificial Societies and Social Simulation*, 16(2):10, 2013.
- [83] Amin Khademi et al. An agent-based model of healthy eating with applications to hypertension. In *Advanced Data Analytics in Health*, pages 43–58. Springer, 2018.
- [84] Donglan Zhang et al. Impact of different policies on unhealthy dietary behaviors in an urban adult population: an agent-based simulation model. *American journal of public health*, 104(7):1217–1222, 2014.
- [85] Philippe J Giabbanelli et al. Modeling the influence of social networks and environment on energy balance and obesity. *Journal of Computational Science*, 3(1-2): 17–27, 2012.

- [86] Tanner Verigin, Philippe J Giabbanelli, and Pål I Davidsen. Supporting a systems approach to healthy weight interventions in british columbia by modeling weight and well-being. In *Proceedings of the 49th Annual Simulation Symposium*, page 9. Society for Computer Simulation International, 2016.
- [87] Saeideh Fallah-Fini et al. Modeling us adult obesity trends: a system dynamics model for estimating energy imbalance gap. *American journal of public health*, 104(7):1230–1239, 2014.
- [88] Vijay K Mago et al. Fuzzy cognitive maps and cellular automata: An evolutionary approach for social systems modelling. *Applied Soft Computing*, 12(12):3771–3784, 2012.
- [89] Philippe J Giabbanelli, Piper J Jackson, and Diane T Finegood. Modelling the joint effect of social determinants and peers on obesity among canadian adults. In *Theories and simulations of complex social systems*, pages 145–160. Springer, 2014.
- [90] Philippe J Giabbanelli and Rik Crutzen. Creating groups with similar expected behavioural response in randomized controlled trials: a fuzzy cognitive map approach. *BMC medical research methodology*, 14(1):130, 2014.
- [91] Venkata Sai Pillutla and Philippe J Giabbanelli. Iterative generation of insight from text collections through mutually reinforcing visualizations and fuzzy cognitive maps. *Applied Soft Computing*, 2019.
- [92] Philippe J Giabbanelli and Piper J Jackson. Using visual analytics to support the integration of expert knowledge in the design of medical models and simulations. *Procedia Computer Science*, 51:755–764, 2015.
- [93] Philippe J Giabbanelli, Andrew A Tawfik, and Vishrant K Gupta. Learning analytics to support teachers assessment of problem solving: A novel application for machine learning and graph algorithms. In *Utilizing Learning Analytics to Support Study Success*, pages 175–199. Springer, 2019.

- [94] Zhao Jianqiang and Gui Xiaolin. Comparison research on text pre-processing methods on twitter sentiment analysis. *IEEE Access*, 5:2870–2879, 2017.
- [95] Tajinder Singh and Madhu Kumari. Role of text pre-processing in twitter sentiment analysis. *Procedia Computer Science*, 89:549–554, 2016.
- [96] Symeon Symeonidis, Dimitrios Effrosynidis, and Avi Arampatzis. A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis. *Expert Systems with Applications*, 2018.
- [97] Kumar HM Keerthi and Harish BS. Classification of short text using various pre-processing techniques: An empirical evaluation. In *Recent Findings in Intelligent Computing Techniques*, pages 19–30. Springer, 2018.
- [98] Melody Barnes. Solving the problem of childhood obesity within a generation. *White House Task Force on Childhood Obesity Report to the President*. Washington, DC, 2010.
- [99] Daghofer D. From weight to well-being: time for shift in paradigms. Technical report, a discussion paper on the inter-relationships among obesity, overweight , 2013.
- [100] Neel Shah, Darryl Willick, and Vijay Mago. A framework for social media data analytics using elasticsearch and kibana. *Wireless Networks*, pages 1–9, 2019.
- [101] Radim Rehurek and Petr Sojka. Software framework for topic modelling with large corpora. In *In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. Citeseer, 2010.
- [102] Kyle Robinson and Vijay Mago. Birds of prey: identifying lexical irregularities in spam on twitter. *Wireless Networks*, pages 1–8, 2018.
- [103] Mannila Sandhu, Philippe J. Giabbanelli, and Vijay K. Mago. From social media to expert reports: the impact of source selection on automatically validating complex conceptual models of obesity. In *Proceedings of the 11th International Conference*

on Social Computing and Social Media, 2019.

- [104] Joseph D Novak. Concept mapping: A useful tool for science education. *Journal of research in science teaching*, 27(10):937–949, 1990.
- [105] Tristan E Johnson, Debra L OConnor, et al. Comparative study of mental model research methods: Relationships among acsmm, smd, mitocar & deep methodologies. 2006.
- [106] Michael J Jacobson and Uri Wilensky. Complex systems in education: Scientific and educational importance and implications for the learning sciences. *The Journal of the learning sciences*, 15(1):11–34, 2006.
- [107] Ruth Zuzovsky and David Mioduser. From concept maps to conceptual models—supporting students’ understanding of complex systems1.
- [108] David Pitt. Mental representation. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2018 edition, 2018.
- [109] Mary Jo Davidson, Laura Dove, and Julie Weltz. Mental models and usability. depaul university, cognitive psychology 404, 1999.
- [110] Bruce Morley. Causality between economic growth and immigration: An ardl bounds testing approach. *Economics Letters*, 90(1):72–76, 2006.
- [111] Yair Wand and Ron Weber. Research commentary: information systems and conceptual modelinga research agenda. *Information systems research*, 13(4):363–376, 2002.
- [112] Doyle KG, Wood JRG, and Wood-Harper AT. Soft systems and systems engineering: on the use of conceptual models in information system development. *Information Systems Journal*, 3(3):187–198, 1993.
- [113] Shadi Moradi Seresht, Olga Ormandjieva, and Samer Sabra. Automatic conceptual

- analysis of user requirements with the requirements engineering assistance diagnostic (read) tool. In *2008 Sixth International Conference on Software Engineering Research, Management and Applications*, pages 133–142. IEEE, 2008.
- [114] Ma ZM, Zhang Wen-Jun, et al. Conceptual design of fuzzy object-oriented databases using extended entity-relationship model. *International Journal of Intelligent Systems*, 16(6):697–711, 2001.
- [115] Jo Anne Earp and Susan T Ennett. Conceptual models for health education research and practice. *Health education research*, 6(2):163–171, 1991.
- [116] Cesar Gonzalez-Perez. A conceptual modelling language for the humanities and social sciences. In *2012 Sixth International Conference on Research Challenges in Information Science (RCIS)*, pages 1–6. IEEE, 2012.
- [117] Jun Chen, Mingrui Shao, et al. Seed v3: Entity-oriented exploratory search in knowledge graphs on tablets. In *Advances in Conceptual Modeling*, pages 7–11, Cham, 2018. Springer International Publishing. ISBN 978-3-030-01391-2.
- [118] Carson Woo, Jiaheng Lu, et al. *Advances in Conceptual Modeling*. Springer, 2018.
- [119] Islay Davies, Peter Green, et al. How do practitioners use conceptual modeling in practice? *Data & Knowledge Engineering*, 58(3):358–380, 2006.
- [120] Cesar Gonzalez-Perez. *What Is Conceptual Modelling?*, pages 3–8. Springer International Publishing, Cham, 2018. ISBN 978-3-319-72652-6. doi: 10.1007/978-3-319-72652-6_1. URL https://doi.org/10.1007/978-3-319-72652-6_1.
- [121] France Légaré, Dawn Stacey, Susie Gagnon, et al. Validating a conceptual model for an inter-professional approach to shared decision making: a mixed methods study. *Journal of evaluation in clinical practice*, 17(4):554–564, 2011.
- [122] Abramova N and Kovriga S. Criterial approach to verification at cognitive mapping of ill-structured situation dynamics. In *Proc. 30th International Conference of the System Dynamics Society.–St. Gallen, Switzerland*, pages 1–23, 2012.

- [123] Nina Abramova and Svetlana Kovriga. The expert approach to verification at cognitive mapping of ill-structured situations. *IFAC Proceedings Volumes*, 44(1): 1997–2002, 2011.
- [124] Robert G Sargent. Verification and validation of simulation models. In *Proceedings of the 2010 Winter Simulation Conference*, pages 166–183. IEEE, 2010.
- [125] Christopher D. Manning, Mihai Surdeanu, et al. The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60, 2014. URL <http://www.aclweb.org/anthology/P/P14/P14-5010>.
- [126] Radim Řehůřek and Petr Sojka. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May 2010. ELRA. <http://is.muni.cz/publication/884893/en>.
- [127] Matthew Hoffman, Francis R Bach, and David M Blei. Online learning for latent dirichlet allocation. In *advances in neural information processing systems*, pages 856–864, 2010.
- [128] Google. Google cloud natural language api. URL <https://cloud.google.com/natural-language/>. Accessed = 2019-04-19.
- [129] Pedregosa F., Varoquaux G., et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [130] Lars Buitinck, Gilles Louppe, et al. API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pages 108–122, 2013.
- [131] George Miller. *WordNet: An electronic lexical database*. MIT press, 1998.
- [132] Steven Bird, Ewan Klein, and Edward Loper. *Natural language processing with Python: analyzing text with the natural language toolkit*. ” O’Reilly Media, Inc.”,

2009.

- [133] Princeton University. Wordnet glossary. URL <https://WordNet.princeton.edu/documentation/wngloss7wn>. Accessed = 2019-04-19.
- [134] Mika V Mäntylä, Daniel Graziotin, and Miikka Kuutila. The evolution of sentiment analysis: a review of research topics, venues, and top cited papers. *Computer Science Review*, 27:5, 2018.
- [135] Alfred S Evans. Causation and disease: the henle-koch postulates revisited. *The Yale journal of biology and medicine*, 49(2):175, 1976.
- [136] Klim McPherson. Causal inference. kenneth j. rothman (ed), epidemiology resources inc., mass., usa, 1988. no. of pages: 207. price: \$21. *Statistics in Medicine*, 10(3):498–499, 1991.
- [137] Bruce G Charlton. Attribution of causation in epidemiology: chain or mosaic? *Journal of Clinical Epidemiology*, 49(1):105–107, 1996.
- [138] Ernst L Wynder. Invited commentary: response to science article, epidemiology faces its limits. *American journal of epidemiology*, 143(8):747–748, 1996.
- [139] Douglas L Weed. On the use of causal criteria. *International Journal of Epidemiology*, 26(6):1137–1141, 1997.
- [140] Vincent Dorie, Jennifer Hill, et al. Automated versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition. *Statistical Science*, 34(1):43–68, 2019.
- [141] Bear F Braumoeller, Giampiero Marra, et al. Flexible causal inference for political science. *Political Analysis*, 26(1):54–71, 2018.
- [142] Jovana Mitrovic, Dino Sejdinovic, and Yee Whye Teh. Causal inference via kernel deviance measures. In *Advances in Neural Information Processing Systems*, pages 6986–6994, 2018.

- [143] Noah Haber, Emily R Smith, et al. Causal language and strength of inference in academic and media articles shared in social media (claims): A systematic review. *PloS one*, 13(5):e0196346, 2018.
- [144] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438, 1969.
- [145] John Geweke. Measurement of linear dependence and feedback between multiple time series. *Journal of the American statistical association*, 77(378):304–313, 1982.
- [146] Mukeshwar Dhamala, Govindan Rangarajan, and Mingzhou Ding. Analyzing information flow in brain networks with nonparametric granger causality. *Neuroimage*, 41(2):354–362, 2008.
- [147] Edwin F Beckenbach and Royal Weller. *Modern Mathematics for the Engineer: First Series*. Courier Corporation, 2013.
- [148] Google Scholar, 2003. URL <https://scholar.google.ca/>.
- [149] JMIR. The journal of medical internet research, 1999. URL <https://www.jmir.org/>.
- [150] BMC. Biomed central, 1999. URL <https://www.biomedcentral.com/>.
- [151] PLOS. The journal of medical internet research, 2000. URL <https://www.plos.org/>.
- [152] IEEE. Institute of electrical and electronics engineers, 1884. URL <https://www.ieee.org/>.
- [153] Twitter. Introduction to tweet json. URL <https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/intro-to-tweet-json>. Accessed = 2019-04-19.

- [154] Susan J Torres and Caryl A Nowson. Relationship between stress, eating behavior, and obesity. *Nutrition*, 23(11-12):887–894, 2007.
- [155] XpdfReader, 1884. URL <http://www.xpdfreader.com/>.
- [156] Skipper Seabold and Josef Perktold. Statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.
- [157] Rolfe Daus Peterson. To tweet or not to tweet: Exploring the determinants of early adoption of twitter by house members in the 111th congress. *The Social Science Journal*, 49(4):430–438, 2012.
- [158] Jacob R Straus, Raymond T Williams, et al. Congressional social media communications: evaluating senate twitter usage. *Online Information Review*, 40(5): 643–659, 2016.
- [159] Vinson C Danielle. *Congress and the Media: Beyond Institutional Power*. Oxford University Press, 2017.
- [160] Kevin M Wagner, Jason Gainous, and Mirya R Holman. I am woman, hear me tweet! gender differences in twitter use among congressional candidates. *Journal of Women, Politics & Policy*, 38(4):430–455, 2017.
- [161] Jennifer Golbeck, Brooke Auxier, et al. Congressional twitter use revisited on the platform’s 10-year anniversary. *Journal of the Association for Information Science and Technology*, 69(8):1067–1070, 2018.
- [162] Annelise Russell. The politics of prioritization: Senators’ attention in 140 characters. In *The Forum*, volume 16, pages 331–356. De Gruyter, 2018.
- [163] John H Parmelee and Shannon L Bichard. *Politics and the Twitter revolution: How tweets influence the relationship between political leaders and the public*. Lexington Books, 2011.
- [164] Tai-Quan Peng, Mengchen Liu, et al. Follower-followee network, communication

- networks, and vote agreement of the us members of congress. *Communication Research*, 43(7):996–1024, 2016.
- [165] Daniel Kreiss. Seizing the moment: The presidential campaigns’ use of twitter during the 2012 electoral cycle. *New media & society*, 18(8):1473–1490, 2016.
- [166] Concha Pérez-Curiel and Pilar Limón Naharro. Political influencers. a study of donald trump’s personal brand on twitter and its impact on the media and users. *Communication & Society*, 32(1), 2019.
- [167] David S Morris. Twitter versus the traditional media: A survey experiment comparing public perceptions of campaign messages in the 2016 us presidential election. *Social Science Computer Review*, 36(4):456–468, 2018.
- [168] Andreas Jungherr. Twitter use in election campaigns: A systematic literature review. *Journal of information technology & politics*, 13(1):72–91, 2016.
- [169] Tsahi Hayat and Tal Samuel-Azran. ”you too, second screeners?” second screeners’ echo chambers during the 2016 us elections primaries. *Journal of Broadcasting & Electronic Media*, 61(2):291–308, 2017.
- [170] Jieun Shin, Lian Jian, et al. Political rumoring on twitter during the 2012 us presidential election: Rumor diffusion and correction. *new media & society*, 19(8): 1214–1235, 2017.
- [171] Sanmitra Bhattacharya, Chao Yang, et al. Perceptions of presidential candidates’ personalities in twitter. *Journal of the Association for Information Science and Technology*, 67(2):249–267, 2016.
- [172] Bryce Jensen Dietrich and Courtney L Juelich. When presidential candidates voice party issues, does twitter listen? *Journal of Elections, Public Opinion and Parties*, 28(2):208–224, 2018.
- [173] Sharon Meraz and Zizi Papacharissi. Networked gatekeeping and networked framing on# egypt. *The international journal of press/politics*, 18(2):138–166, 2013.

- [174] Megan K Maas, Heather L McCauley, et al. "i was grabbed by my pussy and its# notokay": A twitter backlash against donald trump's degrading commentary. *Violence against women*, 24(14):1739–1750, 2018.
- [175] Summer Harlow and Anna Benbrook. How# blacklivesmatter: exploring the role of hip-hop celebrities in constructing racial identity on black twitter. *Information, Communication & Society*, 22(3):352–368, 2019.
- [176] Rashawn Ray, Melissa Brown, et al. Ferguson and the death of michael brown on twitter:# blacklivesmatter,# tcot, and the evolution of collective identities. *Ethnic and racial studies*, 40(11):1797–1813, 2017.
- [177] Leticia Bode and Kajsa E Dalrymple. Politics in 140 characters or less: Campaign communication, network interaction, and political participation on twitter. *Journal of Political Marketing*, 15(4):311–332, 2016.
- [178] Brian L Ott. The age of twitter: Donald j. trump and the politics of debasement. *Critical Studies in Media Communication*, 34(1):59–68, 2017.
- [179] Ashley A Anderson and Heidi E Huntington. Social media, science, and attack discourse: How twitter discussions of climate change use sarcasm and incivility. *Science Communication*, 39(5):598–620, 2017.
- [180] Jenny L Davis, Tony P Love, and Gemma Killen. Seriously funny: The political work of humor on social media. *new media & society*, 20(10):3898–3916, 2018.
- [181] Kareem Darwish. To kavanaugh or not to kavanaugh: That is the polarizing question. *arXiv preprint arXiv:1810.06687*, 2018.
- [182] Aditya Joshi, Pushpak Bhattacharyya, and Mark J Carman. Automatic sarcasm detection: A survey. *ACM Computing Surveys (CSUR)*, 50(5):73, 2017.
- [183] David Bamman and Noah A Smith. Contextualized sarcasm detection on twitter. In *Ninth International AAAI Conference on Web and Social Media*, 2015.

- [184] Mondher Bouazizi and Tomoaki Ohtsuki. Sarcasm detection in twitter:” all your products are incredibly amazing!!!”-are they really? In *2015 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2015.
- [185] Yafeng Ren, Yue Zhang, et al. Context-sensitive twitter sentiment classification using neural network. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.
- [186] Mark Sherman and Lisa Mascaro. Rancorous, partisan start for kavanaugh high court hearing. *Associated Press Newswires*. URL <https://www.apnews.com/23fb3e453a224b3fb0d6575ec926bcd>.
- [187] William Cummings. What we know about the sexual misconduct allegations against brett kavanaugh. *USA Today*. URL <https://www.usatoday.com/story/news/politics/2018/09/24/brett-kavanaugh-allegations-what-we-know-supreme-court/1408474002/>.
- [188] Huyen Le, GR Boynton, et al. Bumps and bruises: Mining presidential campaign announcements on twitter. In *Proceedings of the 28th ACM Conference on Hypertext and Social Media*, pages 215–224. ACM, 2017.
- [189] Huyen T Le, GR Boynton, et al. Revisiting the american voter on twitter. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pages 4507–4519. ACM, 2017.
- [190] Pip Hinman et al. #metoo should supercharge reclaim the night marches. *Green Left Weekly*, (1198):9, 2018.
- [191] Rose McDermott. Political sciences# metoo moment. *Journal of Women, Politics & Policy*, 40(1):148–155, 2019.
- [192] Julie Novkov. The troubled confirmation of justice brett kavanaugh. In *SCOTUS 2018*, pages 125–141. Springer, 2019.
- [193] Barry Sheppard et al. Trump attacks# metoo, attempts supreme court coup.

- Green Left Weekly*, (1198):13, 2018.
- [194] Jennifer Hansen. The hopeful hashtag: Digital feminist publics in the trump era. *American Journal of Theology and Philosophy*, 21(3):217, 2000.
- [195] Hamed Jelodar, Yongli Wang, et al. Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, pages 1–43, 2017.
- [196] Xiaofeng Wang, Matthew S Gerber, and Donald E Brown. Automatic crime prediction using events extracted from twitter posts. In *International conference on social computing, behavioral-cultural modeling, and prediction*, pages 231–238. Springer, 2012.
- [197] Younghoon Kim and Kyuseok Shim. Twilite: A recommendation system for twitter using a probabilistic model based on latent dirichlet allocation. *Information Systems*, 42:59–77, 2014.
- [198] Aniruddha Ghosh and Tony Veale. Fracking sarcasm using neural network. In *Proceedings of the 7th workshop on computational approaches to subjectivity, sentiment and social media analysis*, pages 161–169, 2016.
- [199] Hsuanwei Michelle Chen and Patricia C Franks. Exploring government uses of social media through twitter sentiment analysis. *Journal of Digital Information Management*, 14(5), 2016.
- [200] Chae Won Park and Dae Ryong Seo. Sentiment analysis of twitter corpus related to artificial intelligence assistants. In *2018 5th International Conference on Industrial Engineering and Applications (ICIEA)*, pages 495–498. IEEE, 2018.
- [201] Jyoti Ramteke, Samarth Shah, et al. Election result prediction using twitter sentiment analysis. In *2016 international conference on inventive computation technologies (ICICT)*, volume 1, pages 1–5. IEEE, 2016.
- [202] Lauren M Clohessy. The future of the# metoo movement, 2019. Capstone Thesis.

- [203] Michael Chau and Hsinchun Chen. A machine learning approach to web page filtering using content and structure analysis. *Decision Support Systems*, 44(2): 482–494, 2008.