On Traffic Disruptions: Event Detection from Visual Data and Bayesian Congestion Games

by

Jeffrey Liu

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Civil Engineering and Computation

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2019

Signature redacted

Author .....
Jeffrey Liu
Department of Civil and Environmental Engineering
May 13, 2019

Signature redacted

Certified by .....
Saurabh Amin
Associate Professor of Civil and Environmental Engineering
Thesis Supervisor

Signature redacted

Accepted by .....
Youssef Marzouk
Associate Professor of Aeronautics and Astronautics
Co-Director, Computational Science and Engineering Program

Signature redacted

Accepted by .....
Heidi Nepf
Donald and Martha Harleman Professor of Civil and Environmental Engineering
Chair, Graduate Program Committee

On Traffic Disruptions: Event Detection from Visual Data and Bayesian Congestion Games

by

Jeffrey Liu

Submitted to the Department of Civil and Environmental Engineering
on May 13, 2019, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Civil Engineering and Computation

Abstract

Road traffic is often subject to random disturbances due to weather, incidents, or special events. Effectively detecting and disseminating information about disturbances is a key goal of modern, "smart" infrastructure. Toward this, this dissertation investigates two related questions. First, how can traffic managers better utilize existing traffic cameras to automatically identify traffic disturbances? Second, how can we model different aspects of information—such as human misperception or ignorance of other's information—and their effects on the travelers' route choices?

Part I addresses analyzing unstructured, sequential image data, such as traffic CCTV footage, with a novel, semantics-oriented approach based on natural language and semantic features. The approach extracts structured, human-interpretable "topic signals" from distributions of common object labels, which correspond to physical processes depicted in the footage. Changes and anomalies in these topic signals are used to identify notable events in weather conditions and traffic congestion. This is demonstrated on a new, real-world dataset collected from Boston freeway CCTV footage. In notable event detection, the use of topic signal representation outperforms the use of any individual label signal.

Part II addresses game theoretic modeling of informational effects on travelers' route choices. It considers both access and accuracy of information about the network state, as well as the perception of other's information. It introduces the Subjective Bayesian Congestion Game (BCG), which models a broader set of player beliefs than those allowed by the conventional common prior assumption (Objective BCG). This enables modeling of uncertainty about other's information, such as when one population is unaware of information services. Analytical solutions are provided for a stylized configuration of the Subjective BCG, and a numerical solver is provided for more general configurations. Compared to the Objective BCG, the Subjective BCG has qualitatively distinct solutions and costs, indicating that the perception of other's information significantly affects equilibrium route choices.

Thesis Supervisor: Saurabh Amin
Title: Associate Professor of Civil and Environmental Engineering

Acknowledgments

First and foremost, I'd like to acknowledge my advisor, Prof. Saurabh Amin, for his unwavering support, guidance, and friendship throughout my PhD. This work would not have been possible without his mentorship, and I thank him for all of his time and effort dedicated to his students. In addition, I would like to thank my committee members, Prof. Marzouk and Prof. Zhao, as well as my committee chair, Prof. Osorio, for their expertise and guidance in my research.

Next, I would like to acknowledge my co-authors and my co-workers at MIT, Lincoln Lab, and the Resilient Infrastructure Networks Lab. In particular, I'd like to thank Andrew Weinert for his guidance, discussion, and motivation for the work in Part I; and Manxi Wu, for her analysis, insight, and contributions in Part II. I am extremely grateful for their collaboration and effort in making this work a reality.

I would also like to acknowledge the students that I mentored and taught at MIT. It was a pleasure to work with such motivated and intelligent individuals. In particular, I'd like to recognize David Ogutu for his work, some of which was incorporated into the data collection process, and Jillian Dressler, whose capstone project motivated the work on image retrieval.

Most of all, I'd like to thank my friends and family. In particular, I want to thank my parents and sister, for their consistent support and love throughout my life; my friend and classmate Chi Feng, for his collaboration, knowledge, and friendship; my friend and classmate Renee Bell, for her emotional support and friendship; and my partner, DJ Heebner, for her unbelievable love and kindness.

There are dozens of friends and colleagues whom I do not have the space to mention by name, but have all contributed directly or indirectly to making this work possible; and to them, I express my sincerest gratitude.

The work in this dissertation was performed under the financial assistance of award PSIAP3774 from U.S. Dept. of Commerce, National Institute of Standards and Technology; grants CNS-1239054 and CNS-1453126 from the National Science Foundation; the FM IRG within the Singapore-MIT Alliance for Research and Technology; and the Google Faculty Research Award for "Estimating Social Value of Traffic Information Systems".

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Traffic disruptions, such as accidents, inclement weather, and special events, can adversely affect both the performance and safety of road networks. Current efforts in "smart" infrastructure seek to detect and predict traffic disruptions and their effects in order to communicate this information to infrastructure operators and travelers to inform decision-making. In Fig. 1-1, we illustrate a system of travelers in a transportation network which is subject to random disruptions. Providing information about the state of the network creates a feedback loop (dotted and dashed lines in Fig. 1-1) which affects the route choices of the travelers. Importantly, travelers may have heterogeneous information: they may have different access to information, as well as different beliefs or perceptions about the information that others receive. Thus, the information provided and the quality and travelers' perceptions of that information play a crucial part in determining the behavior and performance of the traffic system.

This dissertation focuses on these informational considerations, and addresses two related problems regarding traffic disruptions. In Part I, we first attend to the problem of detecting anomalous conditions using existing traffic cameras. This is represented by the box labeled "PART I" in Fig. 1-1, and is concerned with the sensing and inference of the network disruption state. In Part II, we address how to model the effects of information distribution on the route choices and costs of travelers; in particular, the aspects of heterogeneous information access, accuracy, and beliefs about others' information. Similarly, this is represented by the box labeled "PART

II" in Fig. 1-1.

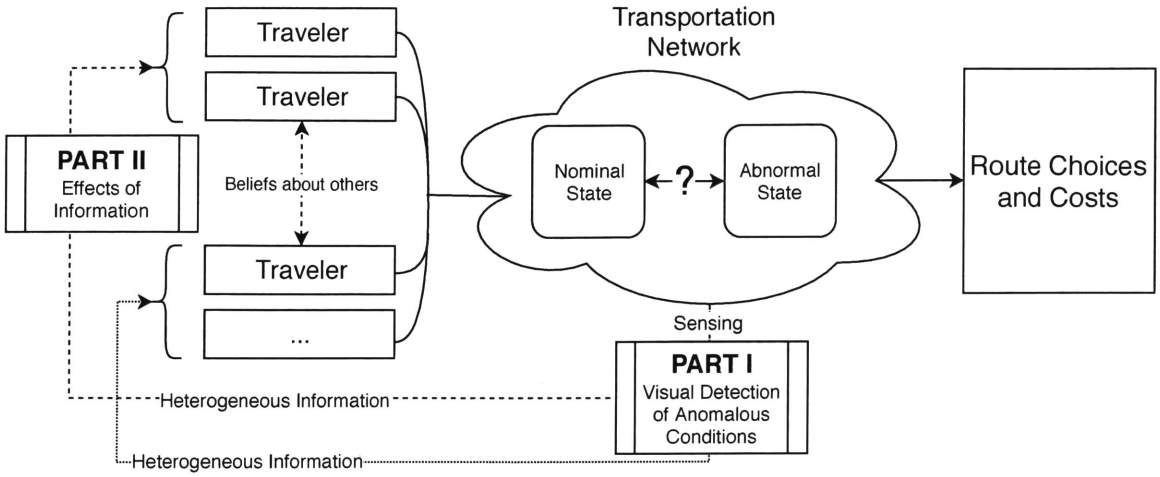


Figure 1-1: Schematic of travelers in a disruption-prone transportation network. The boxes labeled "PART I" and "PART II" correspond to the scope and context of the respective parts of the dissertation.

Both parts are thematically linked by their exploration of the relationship between human decision-makers and traffic information. Part I explores how to use data from existing traffic cameras—which are intuitively understood by human operators, and used to monitor traffic conditions—to automatically detect anomalous traffic conditions. Our techniques seek to preserve human-interpretability for human-in-the-loop control. To accomplish this, we take a natural language, semantics-oriented approach to representing and analyzing the image data. In Part II, we focus on the effects of subjectivity in travelers' beliefs of others' information. This is motivated by the human aspects of misconceptions and irrationalities with respects to information—for example, some travelers may be unaware that other travelers have access to information about traffic conditions, or may incorrectly believe that certain information services are less/more accurate than they actually are.

The behavior of travelers in transportation networks can be studied empirically from data, numerically through simulations, or analytically through models. This dissertation seeks to contribute toward both empirical and analytical approaches. Part I provides an empirical approach, which enables the extraction of signals from a widely-deployed, but underutilized-for-machine-learning data source of traffic cameras. Part

II provides an analytical model of route choices which formally incorporates elements of subjectivity to account for the limitations in human perception and a bounded notion of rationality.

In the following sections, we document the scope, contributions, and related literature of the respective parts in Sec. 1.1 for Part I and Sec. 1.2 for Part II.

1.1 Part I: Semantic Analysis of Traffic Camera Data

Part I addresses the identification of disruption events in traffic camera footage. Traffic cameras were chosen because they are widely deployed, and capture rich, visual information about the infrastructure network. However, the current usage relies on human observers to review the footage from hundreds of cameras, which is labor intensive and inefficient. As such, we believe that the potential of traffic cameras for detecting disruptions is underutilized. We develop a new approach to representing footage using semantic representations of image contents, and apply this approach to identify disruption events within footage.

Closed-Circuit Television (CCTV) traffic cameras are commonly used by Traffic Management Centers (TMCs) for situational awareness, providing 24-hour coverage of rich, intuitive, visual information about driving conditions, infrastructure health, and traffic congestion. Interestingly, most of this footage is kept only for a few hours or days, if at all, before being permanently deleted [1]. In some ways, it is sensible not to store all of the footage: video requires large amounts of disk storage, and most of the time, nothing out of the ordinary is happening. Yet, discarding all of this data also seems like a waste of data from a widely-deployed, flexible sensor, which could be used to improve traffic analytics or diagnostics of infrastructure performance.

Indeed, a small amount of "notable" footage does get saved by operators for personnel training or documentation purposes [1]. However, this process is manual, inefficient, and potentially inconsistent: humans struggle to parse video information

from more than one source at a time [2], and most TMCs have hundreds of cameras which run constantly. It is thus infeasible for human operators to continuously and simultaneously monitor all of the incoming footage. Furthermore, only a few TMCs in the US have written policies regarding what footage is notable enough to be saved [1]. Even for those TMCs that do, the policy's execution is subject to human factors such as subjective interpretation, fatigue, and distraction. Thus, we address in Part I the problem of automatically and consistently identifying notable events from sequential image data, such as traffic CCTV footage.

Previous work on analyzing traffic cameras had focused on pixel-based techniques; the surveys of [3, 4] document techniques for computer vision techniques for urban and freeway traffic applications. Pixel-based techniques utilize the color and geometric configurations of the pixels of the image as features for application-specific classifiers. However, these techniques struggled due to the variability of appearance in objects caused by viewing angle, lighting, and occlusion. Thus, such techniques were only successful in narrowly-defined, well-structured contexts where there is relatively little variation in appearance, such as for license plate recognition [5]. While automated license plate readers are indeed practical, and are often used for automatic tolling and vehicle identification [5], they represent only a narrow application of the broad, general-purpose potential of cameras.

Recently, advancements in deep learning—particularly, convolutional neural networks (CNNs)—as well as in computational power and data availability, have enabled an alternative approach to computer vision. Unlike with traditional computer vision techniques, where the developer would explicitly construct a classifier for a given application, deep learning techniques instead leverage the universal function approximation ability of neural networks [6] to infer the classifier's structure from a massive set of training data. This allows developers to create classifiers for a wide range of object types simply by finding a and labeling sufficient number of training examples for each object type. However, constructing labeled data sets and training the CNNs is laborious and computationally expensive [7]; as such, there are a number of open-source [8, 9, 10, 11] and commercial [12, 13, 14] pre-trained implementations of image

labeling and image classification software.

Pre-trained offerings are generally tailored toward detecting common, everyday objects, and thus may not be suitable for domain-specific applications such as analyzing traffic camera footage. Thus, some researchers have developed their own traffic-specific CNNs to classify traffic footage [15, 16, 17]. While developing application-specific CNNs is certainly viable, we present an alternative approach which leverages the vast amounts of work and data already invested in pre-trained CNNs. Our work analyzes the co-occurrence of labels on images to infer "topics" corresponding to phenomena beyond what the CNN was explicitly trained to recognize. This allows us to detect phenomena such as traffic congestion and winter storms, which the pretrained software was unable to otherwise consistently recognize.

To accomplish this, we develop in Chapter 2 a set of Natural Language Processing (NLP)-inspired, semantic representations of image contents. These semantic representations seek to preserve the intuitive and human-interpretable nature of images. Image contents are represented as Bag-of-Label-Words (BoLW) semantic feature vectors constructed from labels of off-the-shelf image labeling software. These semantic feature vectors are used in the Latent Dirichlet Allocation (LDA) topic model to generate topic representations and *semantic topic signals*. Topics represent distributions of related co-occurring semantic labels, and topic signals are time series that represent how these topics evolve in the footage over time. We identify some topics which correspond to physical processes captured in the footage, such as winter storms and traffic congestion. This approach to semantically representing image contents is novel in both computer vision and intelligent transportation contexts. These methods are evaluated on the Boston Freeway CCTV Camera (BFCC) dataset—a new, public dataset, which we constructed from real-world traffic camera footage.

The semantic representations enable the analysis of image data using existing machine learning and signal processing techniques. In Chapter 3, we address a few analysis tasks using the semantic representations of the images in the BFCC dataset. First, we demonstrate an image retrieval task of finding similar images using the BoLW semantic representation. We show that it is possible to find images corre-

sponding to the same time of day and originating camera with high accuracy using k-Nearest Neighbors. Second, we identify notable events from topic signals using signal processing techniques—in particular, change and anomaly detection applied to the "Wintry Conditions" and "Traffic Congestion" topic signals. For change detection, we implement a mean-shift detector. For the anomaly detection we employ a direct divergence estimation technique based on [18] which does not require parametrically estimating the distributions of the test and reference data.

The semantics-oriented approach introduced in Part I serves as a foundational demonstration of NLP-inspired techniques applied to image data. While this work is built on the well-known NLP Bag-of-Words (BoW) and LDA models, these models are hardly state-of-the-art in NLP. Indeed, we make no claim that our approach is optimal. Instead, it serves as an initial proof-of-concept which establishes potential directions for future work. In particular, framing the analysis of image content labels as an NLP problem unlocks techniques from a rich body of NLP literature for semantic representation, topic modeling, and sequence analysis. Furthermore, extracting time-series signal representations of phenomena in the footage enables techniques from the well-established signal processing and machine learning disciplines. Indeed, the techniques we present are simple, and primarily intended to illustrate the approach. Yet, we are able to achieve respectable performance in validation tasks, outperforming off-the-shelf commercial implementations. That we are able to achieve this performance using these simple, foundational techniques suggests that exploring refinements of our semantics-oriented approach is a promising avenue for future work.

We summarize the contributions for Part I:

- the BFCC dataset of footage from Boston freeway traffic cameras;
- the BoLW semantic label vector representation of image contents;
- the use of the LDA topic model for the extraction of topics and topic signals;
- information retrieval techniques and a semantic similarity score using the BoLW representation;

- and an application of change and anomaly detection on topic and label signals for notable event detection.

These contributions provide a flexible approach to analyzing image data using semantic representations. The techniques are applicable to multiple sensing domains: we demonstrate the detection of notable weather and traffic conditions.

1.1.1 Related Literature

Public traffic CCTV datasets are a relatively recent addition to the literature. We identified two prior, publicly available traffic CCTV datasets: WebCamT [19] and the Car Accident Detection and Prediction (CADP) dataset [20]. WebCamT claimed to be the first publicly available dataset of traffic camera footage [19], and provides detailed annotations for the footage: bounding boxes and labels for vehicle types and weather, as well as vehicle counts and re-identification [19]. The data in WebCamT were collected over four separate, one-hour periods for each day, at a sample rate of one frame per second. The second dataset, CADP, collects and annotates video segments of vehicle crashes from YouTube with vehicle bounding boxes [20]. The videos in CADP are short (a few minutes on average) and intermittent.

The scene labels for the BFCC dataset are generated from a commercial image labeling service powered by deep learning models, such as convolutional neural networks (CNNs). In recent years, the performance of image labeling algorithms have improved to human-comparable error rates in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) benchmark [21]. Though deep learning approaches achieve remarkable performance, they are much more computationally expensive and data-intensive to train than classical approaches [7]. Consequently, many developers and organizations now offer free [8, 9] and commercial [22] pretrained, off-the-shelf object recognition tools and services to detect a wide array of object classes.

CNN-based techniques have been used to recognize traffic congestion directly from images [15, 17, 19]. [15] trains a CNN to recognize different levels of congestion as classes; [17] trains a classifier to segment the image between road and vehicle and

compute the density directly; and [19] takes a similar approach, but also estimates a density map to correct for the distortion effects of perspective and distance in the image. These approaches are indeed performant, but require large amounts of data and computation to train [7]. In comparison, our approach utilizes readily available, off-the-shelf pretrained models. Furthermore, our approach is applicable to a broader set of domains beyond just traffic congestion without retraining. However, we acknowledge that our approach is meant to identify qualitative differences in processes, and is not meant to be a precise estimator of quantities such as vehicle density.

There is a related BoW model in computer vision called Bag-of-Visual-Words (BoVW) [23], where the *visual words* are visual features, such as pixel clusters. In contrast, our "words" are semantic labels—literal textual words. The BoW and BoVW models serve as foundational models for a variety of techniques in NLP and computer vision because they enable linear algebra to be performed on documents and visual contents of images respectively [24, 23]. BoLW seeks to do the same with the semantic contents of images.

LDA is an NLP topic model [25, 26], which are used to find related words in a corpus of documents, as well as characterize the corpus in terms of such topics. There is work in the transportation literature on using topic models to analyze written documents, such as failure reports for railway [27] and aviation applications [28]. Additionally, there has has been work in applying LDA to *visual word* features for dimensionality reduction of image and video data [23, 29]. Semantic features have also been used in [30] to address the cross-view geolocalization problem of finding the satellite image and location corresponding to a ground-view image, showing that semantic representations of images are rich enough to perform tasks as precise as geolocalization using only semantic features.

Anomaly detection is concerned with identifying data that are unlikely to come from a reference distribution [31]. Change detection is a special instance of anomaly detection, where the reference distribution is given by the data in the immediately-preceding time interval. Change detection is well studied in image processing contexts [32]; however, change detection in image processing is typically concerned with detec-

tion of changes in visual elements, such as in shapes, colors, or textures. In contrast, our approach considers changes in the semantic representations of image contents. Where there exist some prior work in using semantic information to contextualize detected visual changes [33], none consider changes in purely-semantic representations.

Our anomaly detection procedure uses divergence measures to quantify the dissimilarity between two distributions of data. We employ a technique which allows us to directly compute the divergence without needing to estimate parametric forms of the respective distributions. These estimation procedures derive from the Kullback–Leibler Importance Estimation Procedure (KLIEP) in [34], and were subsequently extended to other divergence measures in [35, 18]. Direct estimation techniques have been applied to outlier and change point detection generally in [36, 37], but their use in detecting anomalies in semantic representations of sequential image data is novel.

1.2 Part II: Bayesian Congestion Games

Traffic information services have proliferated broadly within the last decade. Through the widespread adoption of smartphones, many travelers now have access to real-time information about the state of the network, including disruptions such as accidents, obstructions, or lane closures. However, not all travelers may have access to the same information; furthermore, they may not know about what information other travelers have. Part II addresses the modeling of route choices of populations of travelers with different access to information about an uncertain state of the network (e.g. traffic accidents). In particular, it focuses on the effects of access and accuracy of information, as well as the perception of other's information. To model this, we introduce the formulation of a population[1] Bayesian Congestion Game (BCG) with *subjective* perceptions of other travelers' information.

Congestion games are used to model the strategic decisions of travelers traversing a network, where the cost of links increase with the number of travelers who utilize the

[1]Population games are a class of games where the "players" are populations of decision makers, and the strategies of each population represent the aggregate result of their individual decisions [38].

link [39]. The well studied [39, 40, 41, 42, 43] classical congestion game formulation is one of complete information, where the cost functions of the network are all common knowledge. These games use the Wardrop Equilibrium [44] as the solution concept. However, these formulations do not capture the potential stochasticity of costs caused by traffic disruptions, nor the heterogeneity and uncertainty of access to information about disruptions.

One approach to incorporating stochasticity are Stochastic User Equilibrium (SUE) models [45, 46] which consider stochasticity through the addition of a random variable which represents individual preferences and perceptions of travel time. While earlier models [45, 46] did not include congestive effects (and thus avoided the need for game theoretic analysis of strategic interactions), later refinements included congestion elements, generally through queuing models [47, 48]. However, the Stochastic User Equilibrium models often use Logit or Probit models, which have the unrealistic feature of assigning some demand on all available routes, regardless of travel time [49]. In addition, the stochasticity of SUE models are typically used to represent the distribution in the preferences of travelers, whereas we would like to capture the stochasticity in the network state caused by disruptions. Finally, SUE models are not game theoretic models, and thus do not incorporate the strategic considerations of travelers and their beliefs about the information that other travelers possess. In order to incorporate both the stochasticity of costs and the strategic considerations of information, we turn to Bayesian games.

Bayesian games are used to model games of incomplete information [50], and thus, we formulate a Bayesian Congestion Game which models stochastic costs and heterogeneous information access and perceptions. While there have been other Bayesian congestion and routing games in the literature [51, 52, 53], our formulation is unique in considering subjective beliefs—i.e. beliefs that are not consistent with the conventional common prior assumption. This allows us to model aspects such as ignorance or misperception of others' information—for example, travelers who are unaware that some travelers have access to an information service—which are otherwise impossible to model under the common prior assumption.

The common prior assumption in Bayesian games [54] assumes that the type distributions, conditional on the game state, of all players are common knowledge. In the context of the Bayesian Congestion Game, this is equivalent to all travelers correctly knowing the accuracy of all other travelers' information; however, this is a limiting assumption. We illustrate with an example why it is necessary to consider the perceptions of others' information: consider a traveler who receives a notification that there is an accident on one of two routes that they can take. If the traveler believes that they are one of the only ones with that information, then they would take the route without the accident (alternate route). However, if the traveler believes that many other people also have that information, then they may not take the alternate route, because everyone else will want to take the alternate route. As a result, the route choice of the traveler depends on what each traveler believes about the other travelers' information. The common prior assumption assumes that the traveler knows the exact likelihood that other travelers received the correct information, which makes it impossible to model cases where travelers are unaware or unsure of the accuracy of the information that other travelers have.

We present the formulation of our Bayesian Congestion Games in Chapter 4. We first present the *Objective BCG*, which is a straightforward Bayesian extension of the complete information congestion game, and adopts the conventional common prior assumption. This formulation is referred to as *Objective* because the beliefs of the players are derived from a common prior which is consistent with the true probability distribution followed by "Nature," the fictitious player who instantiates the game state and player types. We then present the formulation of the *Subjective BCG*, which relaxes the common prior assumption and introduces the notion of *perceived accuracy*—the belief of one traveler population's information accuracy by the other traveler populations. This extension allows the modeling of more nuanced belief structures which do not require traveler's beliefs to be consistent with the distribution used by Nature. We show that the Objective BCG is an edge case of the Subjective BCG, and thus should only used to model systems where common knowledge of all travelers' information accuracies can be justified.

In Chapter 5, we explore the equilibrium route choices and costs of the Subjective BCG on a parallel route network. Note that since the Objective BCG is a special edge case of the Subjective BCG, the equilibrium analysis of the Subjective BCG includes that of the Objective BCG. We show that there are qualitatively distinct solution regimes depending on the probability of incident, the fraction of travelers with information about the network state, and the perception of that information. The effects of subjectivity are captured by a "misperception factor" whose value neatly characterize the qualitative differences in equilibrium strategies between the Objective BCG and the Subjective BCG. We explore how the parameters of information access and perceived accuracy affect the equilibrium costs for each population, as well as the average costs of all travelers (social cost).

Part II provides the following contributions toward the modeling of route choices of travelers in disruption-prone networks:

- formulation of the Objective and Subjective BCGs which model the effects of information about disruptions, as well as the perceived accuracy of that information on the route choices of travelers, respectively with and without adopting the common prior assumption;
- equilibrium solutions of the Objective and Subjective BCGs route choices, and characterization of the solution regimes and the effects of information-related parameters;
- analysis of the population and social costs with respect to solution regimes and information-related parameters;
- and a numerical solution method for the Subjective BCG.

The Subjective BCG provides a game theoretic model which considers strategic decision making of travelers in a disruption-prone network with heterogeneous information and beliefs. The equilibrium results for the Subjective BCG show that the amount of subjectivity in beliefs can meaningfully affect both individual and societal costs on networks, suggesting that how information providers advertise and pose the

accuracy of their information could have a real effect on the travel times in the network. Furthermore, the results show that under certain conditions, it is beneficial for those receiving accurate information to be forthcoming about their access to accurate information; yet in other situations, it is beneficial to hide that they have access to such information. These models can help researchers develop congestion management strategies by managing the distribution and public perception of disruption information, such as through public service announcements and advertisements of information services.

1.2.1 Related Literature

Previous work has identified the need to consider information in modeling travelers' departure time choice ([55], and [56]), and route choices ([57], and [58]). Specifically, [59] identify several empirical potential phenomena that may occur due to travelers having access to traffic information, including concentration, where travelers who receive the same information end up taking the same route, causing congestion; and overreaction, where commuters may incorrectly estimate how others will react to information, leading to oscillations and sub-optimal decisions. Both phenomena are captured in our BCG models: travelers who receive the same information make route choices according to the same strategy, leading to congestion on those routes; and travelers who have incorrect perceptions of the opponent's information will over- or under-estimate the demand of other travelers on each route, resulting in sub-optimal choices.

Our work is built on classical results in congestion games. The existence of a potential function in every congestion game was established by [39] and extended to population games in [43]. Potential games have a unique solution at the extrema of a convex potential function, and were shown to be isomorphic to congestion games [42]. There are also many results regarding the efficiency of equilibrium in classical congestion games [60], [61], [62], [63], and [64].

Moreover, some models of congestion games in asymmetric information environments have been also reported. For example, [65] and [66] showed the existence of

pure Nash equilibrium in Bayesian congestion games with atomic players. [52], and [67] have also studied congestion games with player-specific cost functions. Additionally, [51] studies a congestion game where travelers have different information sets about the available edges (routes), and demonstrates an informational version of the classic Braess paradox. Our work adopts a Bayesian approach to model the information heterogeneity. Particularly, we relax the common prior assumption to address the subjective perception on information accuracy. As a result, the existence, uniqueness, and structural properties of our game's equilibrium cannot be derived by straightforward application of the known results.

More broadly, our work is related to the literature on the value of information. In a classical paper, [68] showed that for a single decision maker, more informative signal always results in higher expected utility. In game-theoretic settings, it is generally difficult to determine whether the value of information in equilibrium for individual players and/or society is positive, zero or negative (see [69], and [70]). The paper [71] shows if the accuracy of the information obtained by an agent is improved, the agent will be better off for sure if other agents are not aware of the change of accuracy. Our focus is complementary in that we analyze the change of strategies and costs of one population when they are believed to have more or less accurate information according to the other population's perception.

There exist related stochastic models of routing, most notably Stochastic User Equilibrium models [45, 46], which model route travel times as random variables corresponding to the perceived travel time for each traveler. While earlier models [45, 46] did not consider congestive effects of other travelers taking the same link, later models incorporated congestion effects, often through queuing models [47, 48]. Our Bayesian Congestion Game explicitly considers congestive effects, as well as the higher-order effects of travelers beliefs of other travelers' behavior, which significantly affects the equilibrium solution. Furthermore, we introduce information services as a mechanism for travelers' beliefs to be correlated. These allow us to model effects of strategic information providers which may be competing with one another to provide better service for their users.

Our formulation of the Subjective Bayesian Congestion Game also relates to broader literature on subjective Bayesian analysis [72] and the common prior [54]. The common prior simplifies the analysis of Bayesian games, allowing for games of incomplete information to be formulated as games of imperfect information [73]. While mathematically convenient, the common prior restricts the set of possibilities that can be modeled. For example, situations where two players have different, incompatible beliefs about a quantity cannot by modeled under the common prior assumption [74]. Yet, public disagreement of opinions and beliefs is commonly observed [54]. Our work relaxes the common prior and assumes a subjective Bayesian interpretation of bounded rationality where players decide based on their own perceptions of others' information accuracy.

Part I

Semantics-Oriented Analysis of Sequential Image Data

Part I focuses on techniques and applications for detecting notable events from images, as highlighted in Figure 1-2. Such detection provides an additional information source for travelers when making their route choices, and also allows transportation network operators to improve awareness and diagnostics of system performance.

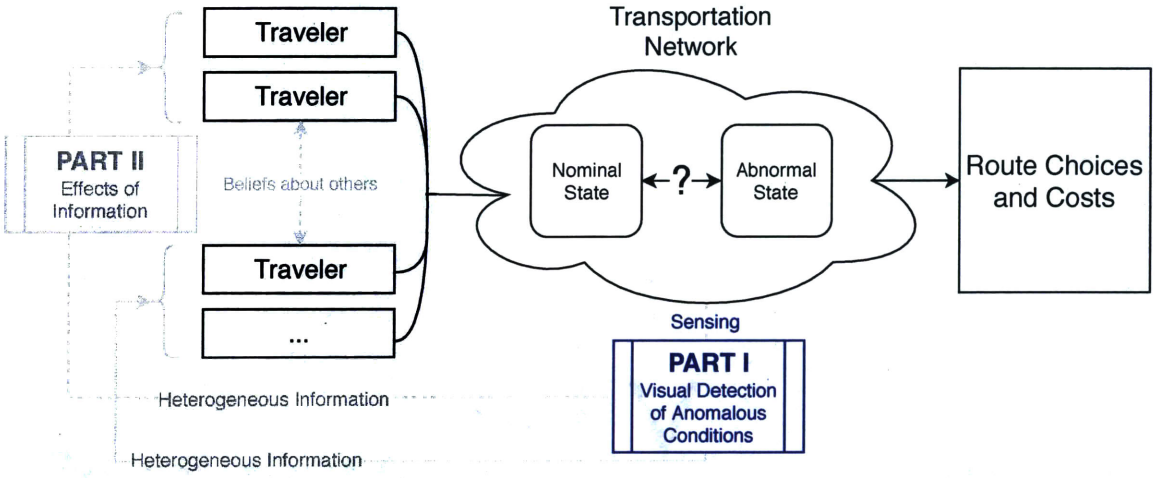


Figure 1-2: Part I addresses the detection of notable events from visual data such as traffic camera images.

We present in Chapter 2 techniques for representing visual image data with semantic features. These techniques are based off of NLP representations of semantic information in documents. Using the semantic representations, we demonstrate a set of analyses and applications, including information retrieval and anomalous event detection in Chapter 3. The work in this part is based primarily off the published and submitted papers [75, 76].

Chapter 2

Semantic Representations

Semantic features capture what objects "are" rather than what they look like—e.g. the word "truck" versus an image of a truck. Semantic features are expressed in natural human language—generally text—which is easily interpretable by human operators. Indeed, the linguistic nature of semantic features inspires us to utilize techniques from Natural Language Processing (NLP), which deals with extracting information from human language, such as unstructured text documents.

In this chapter, we first present the BFCC dataset in Sec. 2.1, which serves as the basis for our empirical work in Part I. The BFCC contains both imagery and machine-generated semantic labels, which we will use for analysis. We then discuss the BoLW semantic feature vector model, which represents the images in a semantic feature space in Sec. 2.2. Finally, we discuss the use of the LDA topic model to identify topics and topic signals in Sec. 2.3.

2.1 Traffic Camera Dataset

2.1.1 CCTV Footage

We collected footage from seven Massachusetts Department of Transportation (MassDOT) freeway CCTV cameras in the Boston metro area. The footage was obtained by scraping the public Mass511 Traveler Information Service website and saved as indi-

ID	Name
1106–1	I-93-ME-Boston-@ exit to HOV-E
1137–1	I-93-NB-Charlestown-@ Zakim South Twr
1296–1	Ramp I-EB-TNL-ramp end 93N x20 c
1413–1	Ramp K-NB-Boston-93N x20 b
1500–1	Ramp K-NB-Boston-93N x20 a
1508–1	Ramp CC-EB-Boston-90E x24C to 93S e
1600–1	Road OHWY-SB-Boston-Leverett Circle

(a) MassDOT camera IDs and names

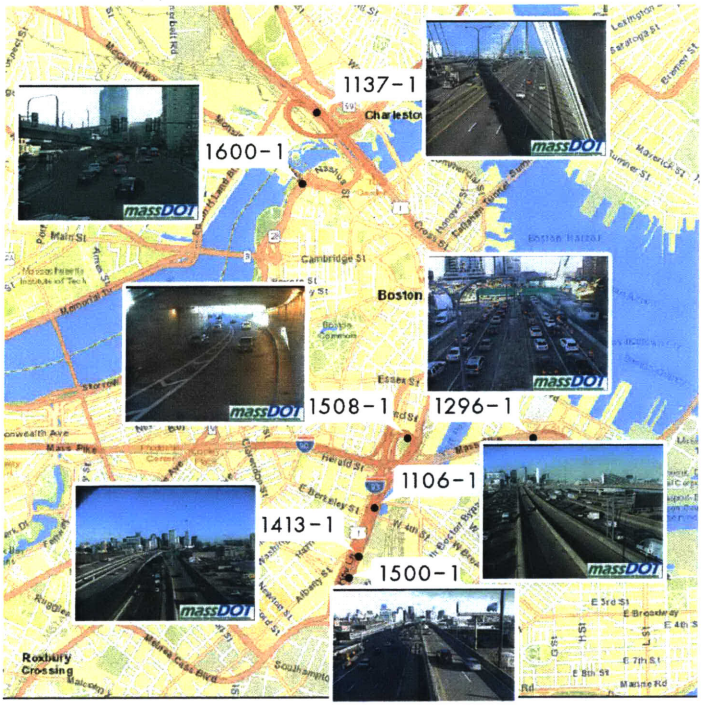


Figure 2-1: Camera locations and sample images. We selected a diverse set of cameras which depicted several different network locations and components, including a bridge (1137–1), underpass (1508–1), intersection (1600–1), HOV lane (1106–1), median (1296–1), and open freeway (1413–1, 1500–1)

vidual frames (also referred to generically as *images*). Each frame has a resolution of 320×240 pixels, and there is a sampling period of roughly 3 minutes between frames, which represents the lower end of camera resolution and frame rate capabilities for typical traffic CCTVs [1].

Fig. 2-1 provides additional details about each camera, including their respective locations, MassDOT-assigned identification numbers and names, and sample images. Each camera is remotely controllable with by the MassDOT TMC operators, with pan/tilt/zoom capabilities. Some of the cameras in the dataset were frequently repo-

sitioned to view alternate perspectives, focus on specific areas of the road, or to avoid obstruction due to snow accumulation.

The data were collected in two phases: phase I consists of the week November 6^{th}–November 12^{th}, 2017; phase II consists of the period between December 17^{th}, 2017–January 31^{st}, 2018. A total of 259830 frames were collected over these two periods. Phase I data serves as an experimental baseline "reference" dataset which was used to establish the system's behavior under nominal conditions. Phase I contained no storms or precipitation, but did include the Veteran's day holiday on Saturday, November 11^{th} (observed on the 10^{th}). The data from phase II serves as the experimental "test" dataset. Notable events that occurred during this phase include: the Christmas and New Year's holidays; several snow storms, including the "bomb cyclone" winter storm of January 2018; and a two-day parking ban imposed by the City of Boston in response to the "bomb cyclone" storm.

Table 2.1 lists the notable events considered for validation tasks in Section 3.2. We considered all snowfall or rainfall events with at least 0.5" of precipitation within a 24 hour period, as reported by the NOAA Global Historical Climatology Network (NOAA-GHCN) daily records [77], as "notable." For holidays and events, we considered the city-imposed parking bans [78] and "major" holidays where retail stores had significantly different hours or were closed[79, 80]. We used the criteria of modified retail hours rather than the federal holiday calendar because not all federal holidays are widely observed, and businesses generally adjust their hours in response to consumer demand. Thus, modified hours are more likely to indicate whether a holiday is widely observed. For this reason, we did not include Veteran's day or Martin Luther King Jr. day as major holidays [81, 82]. Additionally, we did not consider Christmas Eve or New Year's Eve as major events, since they fell on Sundays, and stores in the Boston area observed their typical hours on those dates [79, 80].

Fig. 2-2 illustrates the architecture of the data collection process. The images were scraped off of the public MassDOT Mass511 website [83] using a distributed set of scrapers residing on both local and cloud computers. The image files were stored in a cloud object store database, which were then submitted to a commercial image

Date	Holiday/Event	Rain $> 0.5''$	Snow $> 0.5''$
Dec 23, 2017		✓	
Dec 25, 2017	Christmas		✓
Jan 1, 2018	New Year		
Jan 4, 2018	Parking Ban	✓	✓
Jan 5, 2018	Parking Ban		
Jan 12, 2018		✓	
Jan 13, 2018		✓	
Jan 17, 2018			✓
Jan 23, 2018		✓	
Jan 30, 2018			✓

Table 2.1: List of notable events

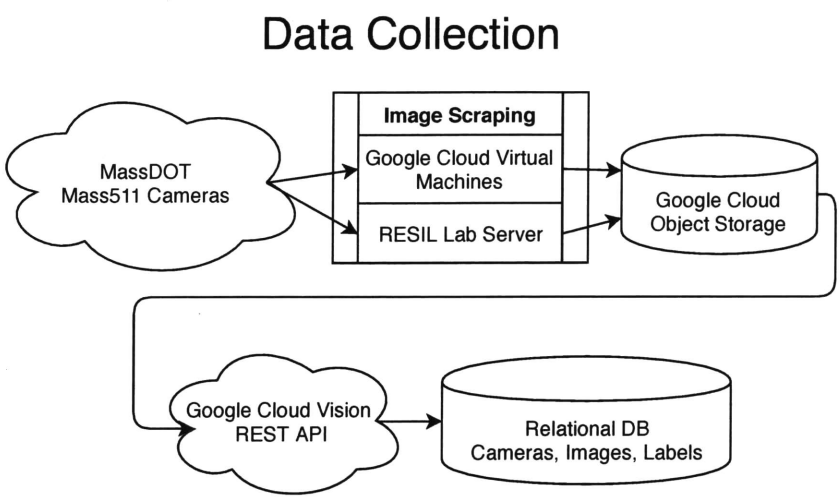


Figure 2-2: Data collection and semantic labeling system architecture

labeling service to obtain semantic labels of the image contents. These labels were stored along with metadata in a relational MySQL database. Next, we will detail this label generation process.

2.1.2 Semantic Feature Labels

We tagged each frame of traffic CCTV footage with *labels* of the image contents using a pretrained, commercially available, common image labeling service: Google Cloud Vision (GCV). GCV offers a number of products, of which we used two: the GCV "label detection" service, referred to as Label Source 1 (LS1), and the GCV "web entity detection" service (resp. LS2). LS1 provides annotations for "broad sets of categories within an image, ranging from modes of transportation to animals," [12],

while LS2 integrates additional information and metadata from the web, such as links and related websites, to detect "web entities"—elements of the Google Knowledge Graph ontology representing real-world entities [12].

Note that our techniques are not exclusive to the GCV services, and can be applied using any image labeling implementation. However, our techniques do assume that the image content recognition problem is a *multi-label* classification problem, where each image can be tagged with multiple labels, as opposed to a *multi-class* problem, where each image is classified into exactly one class [84]. This is because we consider the distribution of labels and their co-occurrence to extract semantic topic signals, which necessitates multiple labels per image.

We chose the GCV commercial implementation because it covers a broad set of categories, is actively maintained and documented, and required the least technical overhead for the user compared to open-source, locally-deployed solutions. In terms of breadth of categories, GCV included labels corresponding to "traffic" and "traffic congestion" in both LS1 and LS2. None of the open source implementations that we examined—models trained on ImageNet [21], COCO [85], and Places [86]—included such labels in their classification set. We use these labels as benchmarks for comparison for the performance of our "Traffic congestion" *topic signal* in Sections 2.3 and 3.2.

In terms of convenience and technical overhead, the commercial implementations required less effort to set up than the open-source ones. The commercial implementations operate as cloud services [22], where the user submits an HTTP POST request with the image, and receives a list of labels in return. This can be done independent of programming language or operating system. In comparison, most pretrained open source implementations required the user to install specific libraries and frameworks in order to run. While this is still easier than training an image labeling model from scratch, it imposes additional technical overhead to the user. For research purposes, the commercial implementations allow for quick annotation of footage and identification of relevant labels from a broad set of object labels.

Fig. 2-3 presents the labels reported by each label service for a sample image taken

from Camera 1137–1 during the "bomb cyclone" blizzard. We refer to the set of all possible labels from the label services as the *vocabulary*. We note that some labels appear in both services, but not necessarily on the same images. For example, "road" appears in both label sources, but for the example image in Fig. 2-3, it is only reported by LS1. Thus, to disambiguate between the labels from each source, we prepend all label text with the respective label source identifier, e.g. "LS1: snow" vs. "LS2: Snow"[1]. This convention would allow additional label sources to be incorporated without ambiguity in future work by prepending the respective labels with "LS3:", "LS4", etc. In this article, if we refer to a label generically without its label source identifier (e.g. the label "snow") we are referring to *both* of the labels from each source (i.e. "LS1: snow" and "LS2: Snow").

(a) Camera 1137–1, 2018-01-04 16:57:52 (UTC)

LS1 labels
snow
infrastructure
mode of transport
lane
winter storm
road
transport
structure
phenomenon
blizzard
highway
freezing
automotive exterior
glass

LS2 labels
Blizzard
Lane
Car
Transport
Snow
Highway
Fog
Glass
Freezing
Massachusetts Department of Transportation

(b) Labels for sample image

Figure 2-3: Fig.(a) shows a sample image taken during the "bomb cyclone," and the table in (b) shows the labels returned by each labeling service

The size of the vocabulary for the dataset between the two label sources is 1389 total labels: 477 from LS1, and 912 from LS2. In general, the labels from LS2 tend to be more specific and contain more named entities than those from LS1. For example, we observed the labels "LS2: BMW," "LS2: BMW 3 Series," and "LS2: 2018 BMW 3 Series Sedan" from LS2, whereas we found only the label "LS1: bmw"

[1]In addition, labels from LS1 were reported by the service in lowercase, whereas those from LS2 were rendered with capitalizations. We preserve this styling.

from LS1. However, LS2 was also prone to including more spurious labels due to word associations: for example, the label "LS2: Blizzard Entertainment" (a software company), appeared occasionally alongside "LS2: Blizzard." However, such spurious labels were rare, and we were able to address this issue in our analysis with a high-pass filter on the labels' empirical document frequency f^j, given by $f^j := n^j / N_{\text{img}}$ where n^j is the number of images in the dataset in which the label j appears, and N_{img} is the total number of images in the dataset.

The cutoff for the high-pass filter is set at $f^j = 10^{-4}$, and was chosen heuristically. We considered that spurious labels may show up once or twice per camera; thus, we set the cutoff at a baseline average rate of three images per camera, which corresponds to a fraction of roughly 0.01% all frames. We also verified that the remaining labels were related to objects and phenomena likely to be observed in traffic footage. In addition, we removed labels from our analysis related to "Massachusetts Department of Transportation," as those labels are likely due to the "massDOT" watermark in the corner of each image, and not the scene content. After filtering, we were left with 620 labels in the vocabulary: 280 from LS1 and 340 from LS2.

2.2 Semantic Feature Vectors: Bag-of-Label-Words

We note that the task of analyzing collections of image labels resembles the classic NLP task of analyzing words in a corpus of text documents [24]. We thus introduce the Bag-of-Label-Words (BoLW) model—adapted from the foundational Bag-of-Words (BoW) [87] NLP model—to represent each image's semantic content in a vector space.

We will consider a set $\mathbf{V}$ of N_{img} images (frames of traffic CCTV footage), indexed by $i \in [1, \ldots, N_{\text{img}}]$. Each image has an originating camera, denoted c_i, and timestamp, denoted t_i. The number of images from a given camera c is denoted N^c_{img}.

2.2.1 BoLW model

The BoLW model for representing image contents as semantic feature vectors is given as follows. Consider a vector space, $\mathcal{L}$, where each dimension corresponds to an

individual label in the label vocabulary. The dimension of $\mathcal{L}$—the number of terms in the label vocabulary—is denoted M_{lab}. A vector in this vector space $\ell \in \mathcal{L}$ represents the labels of image i. The nonzero entries of ℓ_i are equal to unity in the dimensions corresponding to each of the semantic labels for image i. This vector representation is analogous to the Bag-of-Words vector space model of documents in NLP, which represents documents as vectors, where each component corresponds to the number of occurrences of a given word in the document [24]; hence, we refer to our model as Bag-of-Label-Words.

The Bag-of-Label-Words vector model is given as follows:

- A *label word*, λ_j, is defined as a single label in the label vocabulary, indexed by $j \in \{1, \ldots, M_{\text{lab}}\}$. λ_j is a (one-hot) unit-basis vector in $\mathcal{L}$ whose j^{th} component equals one, and all other components equal zero.

- A *bag of label words* associated with image i is a vector $\ell_i \in \mathcal{L}$.

- The total *weight*, w_i, of bag ℓ_i is defined as its L^1-norm: $w_i := \|\ell_i\|_1 = \sum_j |\ell_i^j|$

There is another related BoW model in computer vision, called Bag-of-Visual-Words (BoVW) [88]; however, BoVW uses pixel groupings as its "words," whereas BoLW uses textual, semantic labels as its "words." Generically, these types of "Bag-of-Words"-style models are referred to as "Bag-of-Features" models.

In this dissertation, we consider only labels from*image labeling* classifiers, which provide scene-level annotations of images, and not *object detection*, which also provides counts and locations of each detected object. However, the BoLW model can incorporate object count, simply by adding an additional label word to the bag for each instance of the object detected. For example, if image i had ten "car" and three "truck" labels, its BoLW vector would have the corresponding components $\ell_i^{\text{car}} = 10$ and $\ell_i^{\text{truck}} = 3$.

However, while BoLW can incorporate object count, it does not incorporate location. This is a property shared by all Bag-of-Features models: the configurations of the features—word order in BoW, pixel clusters locations in BoVW, and labeled

object positions in BoLW—are ignored. Instead, the vector representation retains information about the presence and *co-occurrence* of features. This provides invariance to certain transformations of the original data, such as permutations in word order for BoW or rearranging of image elements in BoVW and BoLW. For this reason, Bag-of-Features models tend to be used for capturing general scene-level information (i.e. the *gist* of an image or document) rather than tasks requiring more precise representation (e.g. estimations of geometric quantities such as distance and density). Thus, we focus on detecting categorical, qualitative differences in the scene.

Using BoLW, the semantic content of the footage can be represented in a conventional matrix format. Vertically concatenating the row vectors ℓ_i, ordered by timestamp, for all images of a given camera, generates the $N_{\text{img}}^c \times M_{\text{lab}}$ *image-label* matrix Λ_c. Each row of the image-label matrix corresponds to an image, and each column correspond to a label.[2] This resembles an observation matrix from signal processing: a matrix of N_{img}^c observations of an M_{lab}-dimensional system. Each column can thus be interpreted as the measurements a detector of some quantity; for example, the columns corresponding to the "snow" labels can be interpreted as sensors which detect the presence of snow at that camera's location. The time series for a given label j and camera c are referred to as a *label signal* and are given by:

$$\Lambda_c^j = \{\ell_i^j; \forall i \text{ where } c_i = c\}. \tag{2.1}$$

This representation is remarkable, as it enables the use of existing time series analysis and signal processing methods to analyze complex visual events in videos that are difficult to parameterize in the pixel representation.

2.2.2 Label Reweighting

Note that extremely common labels, such as background elements, do not necessarily contribute much operationally useful information about the image contents. For ex-

[2]The *image-label* matrix is analogous to the *document-term* matrix in NLP, and in general, our usage of the terms "image" and "label" in this dissertation correspond to "document" and "term" respectively in the NLP literature.

ample, labels such as "Road" and "Asphalt" appear extremely frequently in images in the BFCC dataset. While these labels are not incorrect—the images from freeway cameras do indeed contain roads made of asphalt—they are also not particularly informative for TMC operations, as it is expected that most images from a traffic camera contain a road. Thus, we would like to attenuate the weight of labels which occur extremely frequently. this is addressed with the Term Frequency-Inverse Document Frequency (tf-idf) weighting scheme, described below, to rescale each image's label weights based on each label's rarity for each camera.

The tf-idf weighting scheme is a heuristic used in NLP to reweight terms in the BoW vector to account for the natural difference in term prevalence in a language [24]. Terms[3] that are commonly used in a language will be highly represented in any given document, regardless of their relevance to the subject matter of the document. These extremely common terms can end up dominating the weight of a bag if all terms are weighted evenly. Thus, to correct for the effect of these prevalent terms, their weights are scaled inversely to often they appear in all documents. Analogously, labels that appear on nearly every image tend to correspond to static background elements, such as the road and surrounding infrastructure; thus, the same tf-idf reweighting can be used to attenuate these prevalent labels.

The tf-idf weight is computed as the product of its two titular components: the term frequency (tf) and the inverse document frequency (idf) [24]. In NLP usage, the term frequency of a given document and term is given by the number of occurrences of that term within the document; in our case, the term frequency for a given image v and label j is given by the binary variable:

$$\mathrm{tf}(i,j) = \begin{cases} 1 & \text{if image } i \text{ has label } j \\ 0 & \text{otherwise} \end{cases}. \tag{2.2}$$

We use a binary tf term, since we only consider the presence/absence of labels. How-

[3]While in the rest of the dissertation we use the terms "label" and "image" instead of "term" and "document," we preserve the use of "term" and "document" in the explanation of tf-idf in this section due to those words being integral to the tf-idf (*term* frequency-inverse *document* frequency) name.

ever, the term frequency could be used more generally to represent other measures such as object count or number of pixels, if that information is available. This is beyond the scope of this dissertation, but represents a promising refinement for future work.

The inverse document frequency (idf) of a term j is typically computed as the negative logarithm the empirical document frequency: $\text{idf}(j) = -\log(f^j) = \log\left(\frac{N_{\text{img}}}{n^j}\right)$. We use a variant of idf, which we call the *per-camera idf*, which is computed for camera c as $\widetilde{\text{idf}}(j,c) = \log\left(\frac{N^c_{\text{img}}}{n^j}\right)$ where N^c_{img} is the total number of images for camera c. The per-camera idf considers the relative rarity of a label j within the context of the other images from that camera. This is motivated by the fact that the label distributions are different across cameras; for example, the presence of the label "Snow" is more unusual and notable for images from a camera in a tunnel than those from a camera out in the open.

The *idf-weighted image-label* matrix is a rescaling of the image-label matrix where the components of each row (ℓ_i) are given by the per-camera tf-idf values:

$$\ell_i^j = \text{tf}(i,j) \times \widetilde{\text{idf}}(j, c(v_i)). \tag{2.3}$$

The empirical analyses presented in the remainder of this part use the per-camera idf-weighted label data (with the exception of the information retrieval application, as explained in Sec. 3.1); in addition, the data are resampled at a 5-minute interval and linearly interpolated to standardize the sampling rate.

2.2.3 Dataset Structure

We now explore the structure of the data corpus in the BoLW label space. Ideally, the BoLW representation of the corpus would preserve structural relationships between images in the corpus; for example, we would like for images that are similar visually to be close (as measured by some distance metric) in the label space. We first qualitatively assess this by visualizing the data corpus using the t-distributed Stochastic Neighbor Embedding (t-SNE) transformation. In the following chapter,

we will validate this by using semantic similarity to try and find similar images in the corpus.

t-SNE is a dimensionality reduction algorithm typically used to map high dimensional data onto a 2D or 3D space for visualization [89]. The t-SNE technique seeks to preserve the neighborhood relationships between data points—i.e. pairs of points that are nearest neighbors in the high dimensional space should be nearest neighbors in the low dimensional t-SNE space; as a result, local data features such as clusters tend to be preserved. In our case, clusters can be interpreted as groups of data points which are semantically similar to one another. It should be noted that the dimensions of the t-SNE space are arbitrary and relative distances are not preserved in the transformation; thus, the absolute positioning of the data points in the t-SNE space is not inherently meaningful.

We visualize the BoLW representation of the corpus in Figure 2-4 via t-SNE with max iterations=2000 and t-SNE perplexity=100. For clarity of presentation, we plot a random subset of 2×10^4 data points sampled uniformly. Fig. 2-4a colors the data points by camera; we see that the data exhibit some assortativity with respect to camera, where images are likely to be next to others from the same camera. In addition, we observe a strong diurnal effect, illustrated in Fig. 2-4b, where each data point is colored according to its time of day: red for daytime (6am–6pm), and blue for nighttime(6pm–6am).

We partition the data by camera in Fig. 2-4c. We see that each camera appears to have distinct daytime and nighttime clusters. We also note that, in certain cameras (1137–1 and 1508–1), the majority of the images are concentrated in a small number of clusters, whereas in the rest of the cameras, the images are distributed across a larger number of clusters. We found that cameras with fewer changes in camera perspectives tended to have fewer clusters. In the case of camera 1137–1, we only observed four changes in perspective, and in camera 1508–1, we observed none; whereas with the other cameras, we observed upwards of a dozen perspective changes for each, some of which lasted for hours or days before returning to the original angle.

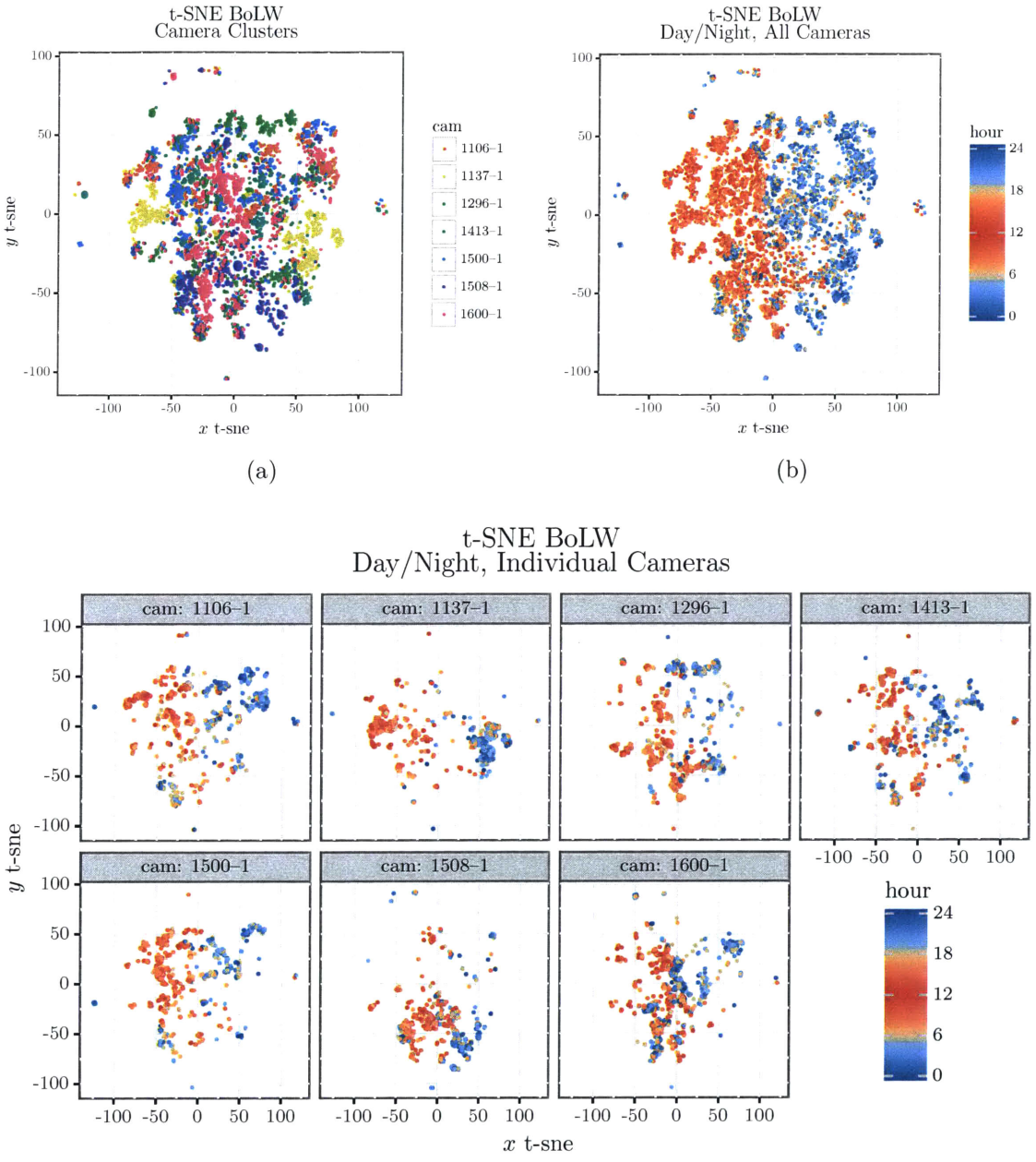
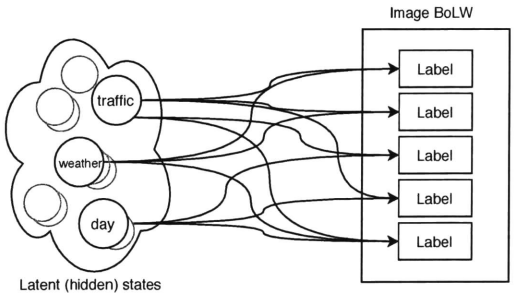


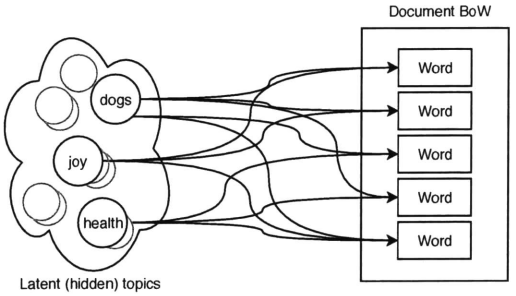
Figure 2-4: Visualization of corpus in BoLW label space via 2D T-SNE projection. The data points in Fig. (a) are colored by camera. The data in Fig. (b) and (c) show the data colored by time of day—red for daytime, and blue for nighttime—for all cameras and each individual camera respectively.

2.3 Topic Signals via Latent Dirichlet Allocation

In this section, we discuss the process of extracting *semantic topic signals* from the BoLW representations of sequential image data. A *topic* represents a distribution of related labels, and can correspond to physical processes. A *semantic topic signal* for a given topic and camera represents a time series of the fraction of the semantic contents of each frame of footage that is related to that topic.



(a) We model each image as an observation of a mixture of stochastic processes which occur on the network.



(b) This is equivalent to the model of documents and topics in NLP probabilistic topic models.

Figure 2-5: Comparison between latent processes model and NLP probabilistic topic models.

The motivation to model physical processes as topics is illustrated in Fig. 2-5. First, certain physical phenomena can be modeled as random processes which generate a mix of objects (and correspondingly, labels) over time. For example, traffic can be seen as a random process which generates cars, trucks, buses, (and their respective labels) etc. at different rates. Weather can be thought of as a random process which generates rain and snow at different rates. Similarly, one can construct processes that

for diurnal lighting cycles and background infrastructure. These processes can thus be modeled as probability distributions over the objects and labels that it generates.

Each image can viewed as an observation of a mix of these *latent* (unobserved) processes. Given a sufficient number of observations (i.e. BoLW vectors of images), one could infer and extract these latent distributions and respective processes captured in the footage. This can be posed as a Bayesian inference problem. Indeed, we recognize that this is equivalent to the problem of probabilistic *topic modeling* in NLP. In particular, we use the Latent Dirichlet Allocation (LDA) topic model of [25]. The identification of topics and processes allow us to represent the footage in much fewer dimensions, and also identifies which labels are most important in detecting each process.

We note that since the *image-label* matrix is an ordinary matrix, other matrix decomposition and dimensionality reduction methods can be also be applied. However, unlike with topic models, not all techniques represent the reduced-dimension data in semantically meaningful formats. For example, the t-SNE transformation we applied for visualization in Fig. 2-4 is a form of dimensionality reduction; yet, the t-SNE dimensions are not inherently meaningful, and cannot be directly used to interpret the semantic information.

Among matrix decompositions with semantically meaningful dimensions, such as singular value decomposition (and the NLP equivalent, Latent Semantic Indexing (LSI) [90]), topic models such as LDA have shown better empirical accuracy in predicting co-occurrence of words, and distinguishing between different usages of the same word [91]. In addition, LSI does not easily incorporate additional data, whereas LDA can be updated in an online manner. For these reasons, we focus on probabilistic topic models.

2.3.1 Latent Dirichlet Allocation Topic Model

Latent Dirichlet Allocation is a hierarchical Bayesian topic model for document generation in NLP [25]. LDA represents documents as random mixtures of topics, denoted θ, where each topic is, in turn, a probability distribution over label words, denoted ϕ.

[26] introduced a commonly used variant that includes an additional Dirichlet prior over on the topic-label distribution ϕ. We adapt this variant of LDA to BoLWs below, and visualize it in Fig. 2-6.

Each bag of label words ℓ_i in a dataset $\mathbf{V}$ is generated by the LDA model with the procedure described in Algorithm 1. The target weights $\overline{w}_i$ are set exogenously based on the empirical bag weights of $\mathbf{V}$; in addition, the target weights are rounded to the nearest integer. [4]

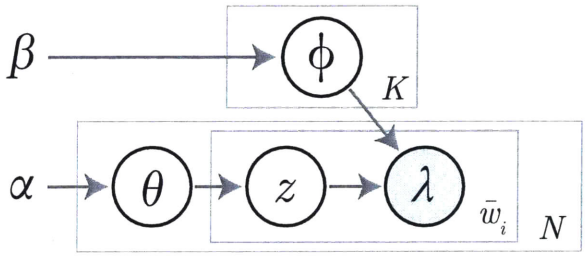


Figure 2-6: Graphical representation of the LDA model structure. Each of the boxes (plates) represent a repeated component; the variable in the lower right hand corner of each plate indicates the number of copies. The outer plates represent each bag of label words in the dataset, and the inner plate represents each label word added to the bag. Grey-filled circles represent observed variables, whereas white-filled circles represent latent variables.

Algorithm 1: LDA BoLW generation procedure

input : Target weight, $\overline{w}_i$, of bag ℓ_i
image-topic prior hyperparamater α
topic-label prior hyperparamater β
output: bag of label words ℓ_i
Initialize $\ell_i = \mathbf{0}$
Draw $\theta \sim \text{Dirichlet}(\alpha)$
Draw $\phi \sim \text{Dirichlet}(\beta)$
while $w_i < \overline{w}_i$ **do**
  Draw topic $z \sim \text{Multinomial}(\theta_i)$
  Draw a label word $\lambda \sim \text{Multinomial}(\phi_z)$
  Add λ to the bag: $\ell_i = \ell_i + \lambda$
end

A topic is denoted $z \in \{1, \ldots, M_{\text{top}}\}$, where M_{top} is the total number of topics, set exogenously. The *topic-label* distribution is denoted ϕ, and is drawn from a Dirichlet

[4] Rounding can be done without loss of generality; an arbitrary p additional significant digits can be preserved by first scaling all elements in the bag by by 10^p.

distribution characterized by M_{lab}-dimensional hyperparameter β. The distribution of labels that constitute a given topic, z, is denoted $\phi_z = \phi(\lambda|z)$. The *image-topic* distribution is denoted θ and characterized by the $M_{\text{top}}-$dimensional hyperparameter α. The marginal distribution for a given topic is denoted $\theta^z(i) = \theta(i|z)$.

We define the *semantic topic signal* of a given topic z and camera c as the time series:

$$\Theta_c^z := \{\theta^z(i); \forall i \text{ where } c_i = c\}. \tag{2.4}$$

The topic signal represents the proportion of the camera's semantic weight which corresponds to topic z over time. Increases/decreases in this signal correspond to a respective increase/decrease in the fraction of labels related to the topic. Each individual topic signal can be analyzed as a univariate time series, and combinations of topic signals can be analyzed jointly.

To fit the model, we want to find the most likely (i.e. maximum posterior probability) values for the *image-topic* distribution, θ, and *topic-word* distribution, ϕ, given the hyperparameters and dataset $\mathbf{V}$. This is done using the online variational Bayes algorithm presented in [92]. We assume symmetric priors on θ and ϕ with constant hyperparameter values $\alpha = 50/M_{\text{top}}$ and $\beta = 0.1$ based on [26].

2.3.2 Choosing Appropriate Number of LDA Topics

The number of topics in the LDA model, M_{top}, is specified exogenously—i.e. it is not inferred by the model. A larger value of M_{top} can account for more distinct processes, at the expense of increasing model complexity. We use the perplexity metric to choose the appropriate number of topics for the model. Perplexity is an entropy-based metric for assessing how well a probability model predicts an unseen set of test images, $\mathbf{V}_{\text{test}}$ [25], given by:

$$\text{Perp}(\mathbf{V}_{\text{test}}) := \exp\left(-\frac{\sum_{i \in \mathbf{V}_{\text{test}}} \log(p(\ell_i))}{\sum_{i \in \mathbf{V}_{\text{test}}} w_i}\right).$$

where $p(\ell_i)$ is the likelihood of the model generating the label vector ℓ_i. In our case, we use $p(\ell_i) = p(\ell_i|\beta, \phi)$, the conditional likelihood of observing ℓ_i from a LDA model given the hyperparameter β and fitted topic-label distribution ϕ:

$$p(\ell_i|\alpha, \phi) = \int p(\theta^i|\alpha) \left(\sum_{j=1}^{\overline{w}_i} p(\lambda_j|z_j, \phi) p(z_j|\theta^i) \right) d\theta^i.$$

We select the appropriate number of topics, denoted M^*_{top}, in a manner similar to [93]. Since a lower perplexity score indicates a better fit of the model to the data, we increase M_{top} until we no longer see an appreciable decrease in perplexity. Let $\text{Perp}_{M_{\text{top}}}(\mathbf{V}_{\text{test}})$ denote the perplexity of a holdout dataset $\mathbf{V}_{\text{test}}$ for an LDA model with M_{top} topics. The data was partitioned at random into an 80/20 train/test split. Several LDA models were fit over a range of M_{top}, and we compute the Rate of Perplexity Change—a finite difference approximation of the slope with respect to M_{top}—as:

$$\text{RPC}(K) := \frac{\text{Perp}_{M_{\text{top}}}(\mathbf{V}_{\text{test}}) - \text{Perp}_{M_{\text{top}} - \Delta M_{\text{top}}}(\mathbf{V}_{\text{test}})}{\Delta M_{\text{top}}}.$$

Figure 2-7 shows the rate of perplexity change versus the number of topics; the error bars represent the standard deviation of 50 Monte Carlo resamplings, with random train/test data partitions for each resampling. We select the smallest M_{top} within one standard deviation from zero as the number of topics, $M^*_{\text{top}} = 20$.

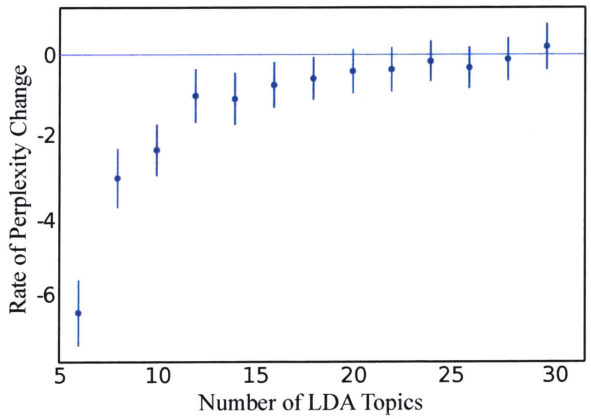


Figure 2-7: Rate of perplexity change vs. Number of topics; error bars show standard deviation from Monte Carlo samples

Topic 1: Wintry Conditions	**Topic 8:** Nighttime Street Lights	**Topic 9:** Intersection	**Topic 11:** Traffic Congestion	**Topic 12:** Error
LS1: snow	LS2: Street	LS2: Intersection	LS1: vehicle	LS1: white
LS2: Snow	LS1: street light	LS1: intersection	LS2: Vehicle	LS1: material
LS2: Phenomenon	LS2: Lighting	LS1: skyway	LS1: motor vehicle	LS2: Webcam
LS1: geological phenomenon	LS2: Street light	LS1: urban area	LS2: Motor vehicle	LS1: circle
LS1: phenomenon	LS1: night	LS2: Urban area	LS1: automotive exterior	LS1: technology

(a) Sample of LDA topics, and their respective highest probability labels in descending order

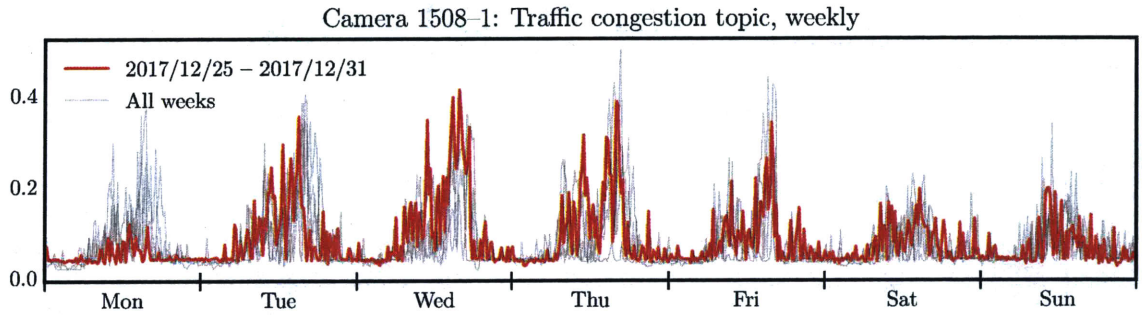
(b) Error message that is shown when a live feed for a camera is unavailable

Figure 2-8: Selected LDA topics (a) and unavailable feed error message (b)

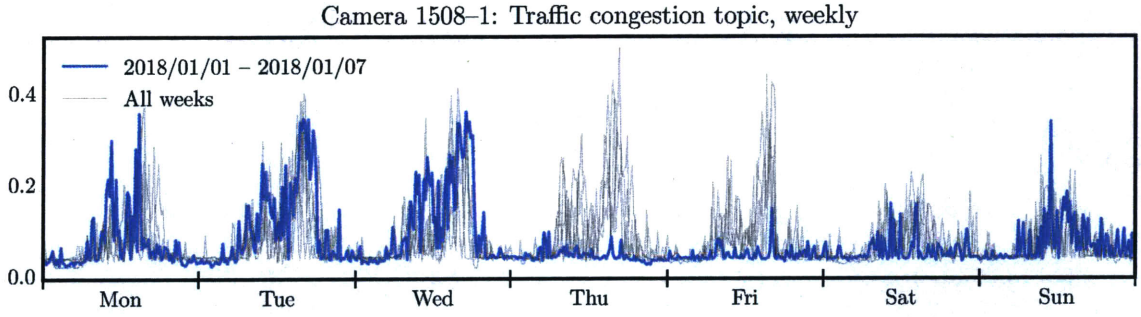
2.3.3 Selected Topics and Signals

We now discuss a few selected semantic topics and topic signals. These results come from an LDA model with $M^*_{\text{top}} = 20$ topics, fit on the entire dataset, weighted using the per-camera tf-idf scheme. Fig. 2-8a presents a handful of representative topics and their five highest-probability labels. Recall that the tf-idf scheme reweights labels relative to their average appearance frequency. Without this reweighting, the highest probability labels of each topic would be dominated by the most common labels of "road" and "asphalt".

The LDA model represents each image in the dataset as a mixture of the 20 LDA topics, where the fraction of each topic corresponds to the fraction of the image's semantic weight associated with that topic. The *semantic topic signal* represents that fraction, for a given topic and camera, as a function of time. We adopt the common LDA practice of naming each topic, *a posteriori* based on domain knowledge and un-



(a) Week of Christmas (Mon)

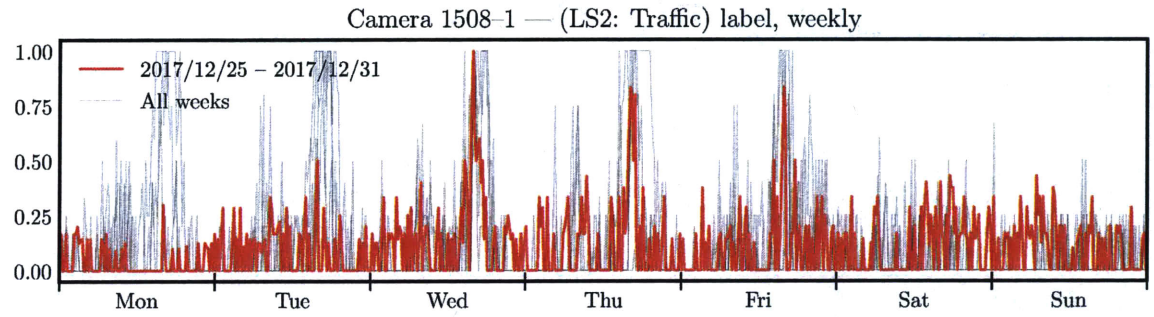


(b) Week of New Year's (Mon) and "Bomb cyclone" (Thu–Fri)

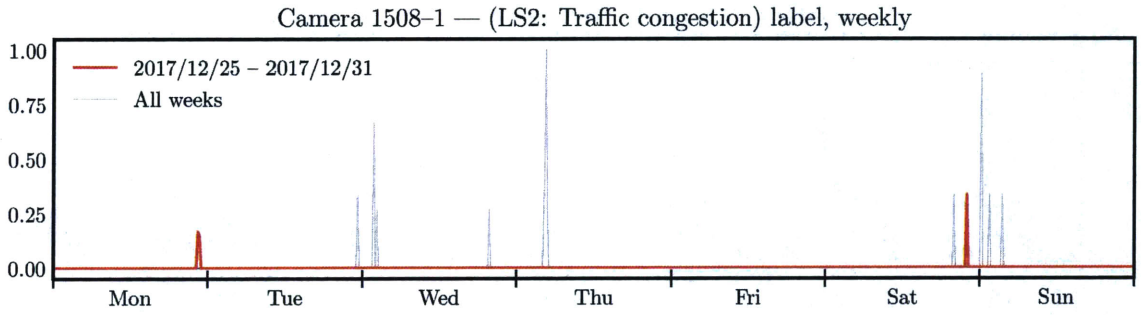
Figure 2-9: Camera 1508–1 "traffic congestion" topic signals, with weeks superimposed. Fig. (a) highlights the week of Christmas; Fig. (b) highlights the week of New Year's and the "Bomb cyclone" storm.

derstanding of the labels. We associate topics with physical *processes* which generate labels corresponding to elements related to that process. We find that the topics cover categories of processes including environmental/weather phenomena, diurnal cycles, infrastructure elements, traffic, and error messages.

It is interesting to note that the semantic meanings are not considered in the LDA topic inference process, yet the statistical LDA process seems to aggregate semantically similar labels into topics. The exception here is in "Topic 12: Error," which has labels that are seemingly unrelated to traffic CCTV footage, as well as to one another. However, once we realize that Topic 12 appears only for the image shown in Fig. 2-8b, which is given by the Mass511 web server when the feed is temporarily down, the relation becomes clear. This semantic similarity and ease of interpretation is an intended feature of the topic model approach, which aims to retain the intuitive parsability of image data. Furthermore, this demonstrates a useful side effect of using the LDA representation: automatic identification of frames with recurring error



(a) "LS2: Traffic" label



(b) "LS2: Traffic congestion" label

Figure 2-10: Camera 1508–1 individual label signals vs. time, with weeks superimposed. Fig. (a) shows the signal for the label "LS2: Traffic"; Fig. (b) shows the signal for the label "LS2: Traffic congestion".

messages.

We find that "Topic 1: Wintry Conditions" corresponds to winter storm events. Unsurprisingly, the top labels include "snow" from both label sources. However, unexpectedly, the next three labels are variations of "phenomenon" and "geological phenomenon", which we did not expect *a priori*, to correspond to winter storm events. However, we find in notable event detection (presented in the next section), that the topic performs better in validation than using naively only "snow" and/or "rain" labels. This suggests that the "phenomenon" and "geological phenomenon" labels provide useful information toward detecting winter storm events. This demonstrates another benefit of the semantic topic representation: the ability to discover and identify labels that are related to quantities of interest.

We now examine the "Topic 11: Traffic congestion" topic signal to qualitatively gauge traffic congestion patterns from the footage. Fig. 2-9 presents the "Traffic congestion" topic signal for camera 1508–1, with each week superimposed. The data

points are plotted at 15-minute intervals (downsampled using mean value). Camera 1508–1 is selected for its location in an underpass, which protects it from atmospheric occlusion due to rain or snow. In addition, for the duration of the data collection period, the camera angle was not manipulated by operators. We observe a diurnal pattern of more traffic congestion during the day, as well as a weekly pattern of lower congestion on the weekends. Furthermore, we see more congestion during the evening rush hours than in the morning, as is expected from typical urban commuting patterns, since the camera is located on a ramp leading out of Boston. These phenomena match our expectations of traffic flow at this location.

We also highlight two weeks to show the "Traffic congestion" topic signal's sensitivity to holidays and major storms. Fig. 2-9a highlights the week of Christmas, which shows a clear reduction in traffic congestion on Christmas day. Fig. 2-9b highlights the week of New Year's and the "Bomb cyclone" winter storm, which occurred on Monday and Thursday–Friday respectively. We see that the New Year's reduction in traffic was not as dramatic compared to that of Christmas; this is consistent with expectations, as in the United States, nearly all businesses and organizations are closed on Christmas, but many businesses are open on New Year's, albeit often with reduced hours [80].

While the effect of New Year's was relatively mild, the "Bomb cyclone" had a much more stark impact on traffic, reducing it to effectively zero for much of Thursday and Friday. Though the storm itself did not reach Boston until just past midnight on Friday morning, traffic was virtually nonexistent for all of Thursday. This is likely due to the City of Boston imposing a parking ban, which was in place from 7 a.m. Thursday–5 p.m. Friday [94, 78]. We see a small uptick in traffic around 5 p.m. on Friday right when the parking ban is lifted. Again, Topic 11's signal behavior is in line with our expectations for these traffic disturbance events.

For comparison, we provide similar weekly graphs constructed using individual labels in Fig. 2-10. Like with the topic signal, we plot the label signals at a 15-minute interval using mean-value downsampling. Fig. 2-10a shows the "LS2: Traffic" label. We observe that there is a significant background level of noise, with roughly one in

every five images being tagged with "LS2: Traffic" throughout, including at night. The label signal does seem to capture the general phenomena: the afternoon peak in traffic, and reduced traffic on weekends and holidays. However, it is difficult to distinguish between the smaller variations in this signal, such as differences between weekend daytime and nighttime traffic.

Fig. 2-10b shows the label signal for "LS2: Traffic congestion". It is clear that this label signal fails to consistently detect traffic congestion, since the label appears only a handful of times, and generally outside of high-traffic rush hours. The label signal does not capture any of the expected diurnal, weekly, or event-related patterns. The plots for "LS1: traffic" and "LS1: traffic congestion" were omitted, since the former looks similar to its LS2 counterpart, and the latter does not appear at all on any images for camera 1508–1.

These graphs suggest that the topic signal provides a better representation of a "traffic congestion" process which captures more of the phenomena we expect to observe than label signals do. The following section validates this by comparing the performance of using topic signals versus label signals for detecting notable events.

Chapter 3

Semantic Analysis Applications

3.1 Information Retrieval

The BoW model is often used in NLP information retrieval applications, which are concerned with finding documents which are related to a given query document [24]. In this section, we demonstrate that the BoLW representation can be used in the same way to identify images that are semantically similar to a query image. For the analysis in this section, we considered a random subset of 15000 images and their respective BoLW representations from the BFCC, partitioned into an 80/20 split. The 80% subset represents the set of retrievable images, whereas the 20% set represents the set of query images.

We identify semantically similar images using the angular distance metric (defined in Sec. 3.1.1) computed on the BoLW[1] representations of the query image and retrievable images. We use a nearest-neighbors approach in finding related images.

We consider whether the retrieved images are "related" to the query image using a few metrics: same camera and same time of day. We evaluate the results on the "top-k" retrieved image performance: if the majority of the top k retrieved images

[1]We do not use the per-camera tf-idf weighted data for the BoLW representations in this section. The per-camera tf-idf weights implicitly encode information about each camera through the per-camera idf term, because all images from the same camera receive the same scaling. As such, if the query image's BoLW representation were already weighted using per-camera tf-idf, it would be biased to be closer to images from the original camera. Thus, to avoid this bias, we do not scale any of the BoLW vectors for the information retrieval task.

have the same camera or time of day (within $\pm$ 2 hour window) as the query image, then we consider those results a successful query in the respective camera/time of day metric.

We first perform the query using exhaustive search on all 12000 retrievable images, and report the results in Sec. 3.1.2. While the results of the exhaustive search are relatively accurate, they are computationally intensive. We are able to achieve a significant computational improvement using a cluster-based indexing technique to reduce the runtime, with only minor accuracy tradeoffs, discussed in Sec. 3.1.3.

3.1.1 Semantic Similarity and Performance Evaluation

We measure the semantic similarity between two BoLW vectors using the cosine similarity score. Cosine similarity measures the alignment between two vectors, and is commonly used in information retrieval contexts for high dimensional data [24]. The cosine similarity between two BoLW vectors ℓ, ℓ' is given by:

$$\cos\ \mathrm{sim}(\ell, \ell') = \frac{\ell \cdot \ell'}{\|\ell\| \|\ell'\|}$$

where $\cdot$ denotes the dot product and $\|\ell\|$ denotes the Euclidean norm of ℓ. Since the elements of BoLW vectors are non-negative, the cosine similarity is bounded on $[0, 1]$, where 0 represents complete dissimilarity (orthogonality) and 1 represents maximum similarity (alignment) between the two vectors.

Since we are using a nearest-neighbors approach to finding similar images, it is convenient to represent the cosine similarity as a *distance* such that larger values represent greater dissimilarity between vectors. To do this, we use the angular distance [24], simply defined as:

$$\mathrm{ang\ dist}(\ell, \ell') = 1 - \cos\ \mathrm{sim}(\ell, \ell').$$

Since the evaluation of whether an image is "related" to another is somewhat

subjective, we instead use other properties as proxy metrics for "relatedness." In particular, we consider two metrics for result relevance: same camera and same time-of-day. We recognize that these properties are not sufficient to establish whether a retrieved image is relevant to the query image, but we argue that they are necessary; furthermore, we have the benefit of having the ground truth values of these quantities. Future work would be needed to further evaluate the performance of information retrieval using the BoLW semantic representations.

The "same-camera" metric between two images, i and i', is defined as a boolean function whose value is "true" if and only if the two images originated from the same camera. Similarily, we define the "same-time-of-day" metric between two images as a boolean function whose value is "true" if and only if the timestamps of the two images differ by less than two hours.

$$\text{same-cam}(i,i') = \begin{cases} \text{True} & \text{if } c_i = c_{i'} \\ \text{False} & \text{otherwise} \end{cases}$$

$$\text{same-time}(i,i') = \begin{cases} \text{True} & \text{if } |t_i - t_{i'}| \leq 2 \text{ hours} \\ \text{False} & \text{otherwise} \end{cases}$$

Individual image retrieval results are evaluated using a top-k performance; that is, the image retrieval results considered correct for the same-camera or same-time-of-day metric if the majority of the top-k images satisfy that metric. We explore the performance for different values of $k \in \{1,3,5\}$. The overall performance is evaluated using an accuracy score for each metric, which is defined as the fraction of query images which had correct top-k results for the respective metric.

3.1.2 Exhaustive Search

We first evaluated the image retrieval using exhaustive search to find the top-k nearest neighbors on all 12000 images in the retrievable image set. The accuracy for each metric, along with the runtimes are listed in Table 3.1. For reference, the null classifier

	same-cam accuracy	same-time accuracy	runtime
k=1	80%	84%	2 hours
k=3	86%	92%	2 hours
k=5	87%	92%	2 hours

Table 3.1: Image Retrieval Performance, Exhaustive Search

accuracies (guessing uniform at random) are 14% and 17% for the same-cam and same-time metrics respectively. We see that in all top-k evaluations, the nearest-neighbor retrieval approach performs much better than the null classifier and achieves reasonable results.

		Retrieved Image Camera						
	k=3	1106--1	1137--1	1296--1	1413--1	1500--1	1508--1	1600--1
Query	1106--1	0.122	0.001	0.002	0.007	0.008	0.001	0.002
Image	1137--1	0.001	0.12	0.003	0.001	0.004	0.002	0.004
Camera	1296--1	0.002	0.001	0.124	0.001	0.003	0.002	0.004
	1413--1	0.008	0	0.001	0.125	0.012	0.001	0.004
	1500--1	0.008	0.005	0.002	0.013	0.118	0.001	0.004
	1508--1	0.002	0.002	0.001	0.001	0.001	0.135	0.003
	1600--1	0.002	0.004	0.003	0.004	0.004	0.003	0.118

Figure 3-1: Confusion matrix of retrieved images for $k = 3$, same-camera metric. Computation times are given for a quad-core intel laptop processor.

We briefly note some observations regarding the confusion matrix for the same-camera metric, shown in Fig. 3-1 for $k = 3$. The confusion matrix shows how the camera classifications are distributed between query and retrieved images. A disproportionate number of misclassifications were between cameras 1106–1, 1413–1, and 1500–1. We note that all three cameras are located on the same freeway (see Fig. 2-1), have similar vantage points, and are all pointed north. As such, their images look visually similar—the fact that their images were more likely to be confused with one another than with other cameras suggests that this element of visual similarity is reflected in the semantic representation.

We also note that camera 1508–1 had comparatively less misclassification relative to other cameras. This is likely due to its unique positioning in an underpass, as well as its total lack of camera angle changes. As a result, most of the images from

1508–1 depict a similarly consistent scene from day to day, and the images have less dispersion in the BoLW space compared to other cameras. This decreases the likelihood of images being adjacent to those from other cameras.

On the other hand, camera 1600–1 has a relatively high misclassification compared to other cameras. This is likely due to the high number of camera angle changes observed in this camera. Operators would frequently change the perspective of this camera, often for short durations. Consequently, there's a larger number of small clusters of images corresponding to these perspective changes which have relatively small number of neighbors. As such, some of the nearest neighbors end up being images from other cameras, leading to misclassification.

3.1.3 Cluster-Based Indexing

While the exhaustive search performs quite well in terms of accuracy, the computational expense is high, since we must compute the angular distance between every query image and every retrievable image. We present a cluster-based indexing approach to reduce the additional queries by reducing the number of candidate retrievable images to consider.

The clustering approach is motivated by the observed cluster structures in the t-SNE projection (Fig. 2-4). We see that each camera appears to have some distinct local clustering, which we can use to reduce the number of candidate retrievable images to consider. If we can establish that the query image is within a cluster, we can restrict the set of candidate images to only those within the cluster. The query process is thus broken into two steps: first, identify which clusters are near the query image, and second, search exhaustively through only the images in the clusters near the query image.

This process reduces the number of images needed to be considered from the full 12000 to a few hundred, resulting in a massive decrease in runtime from hours to minutes. However, there is an one-time upfront computational cost of about three hours incurred in order to find the cluster membership of all of the retrievable images. We use hierarchical clustering [95] to find the cluster membership. We chose hierarchical

	same-cam accuracy	same-time accuracy	runtime
k=1	72%	82%	0.5 minutes
k=3	84%	91%	3 minutes
k=5	87%	92%	6 minutes

Table 3.2: Image Retrieval Performance, Cluster-Based Indexing

clustering because it allows for non-convex clusters, as well as heterogenous cluster sizes, which we observe in the data. We chose 517 clusters based on the L Method of [96].

We constructed an index set of 1528 retrievable images, consisting of up to 5 points (medoid and up to 4 random images) from each of the 517 clusters. In the first step, we find the k-nearest-neighbors out of the index set. Then we select the clusters to which those k-nearest index neighbors belong to, and exhaustively search through those clusters to find the k-nearest images to retrieve in the second step. The performance for this cluster-based indexing technique is shown in Table 3.2, using the same 2000 query images. We see that the accuracy is a little worse for $k = 1,3$, but comparable to the exhaustive search for $k = 5$.

The runtime for the cluster-based indexing is orders of magnitude quicker than the exhaustive search, completing the full query in a manner of minutes rather than hours. The additional upfront cost of hierarchical clustering is recouped entirely if one performs more than two batch queries against the retrievable image set.

These results indicate that semantic representations, such as BoLW can be used for information retrieval tasks. Furthermore, we can exploit the structure of the image data to index the data based on clustering to reduce the computational complexity of searching through the images. Future work is needed to further evaluate the quality of the query results, and there is significant room for improvement by considering additional features, such as visual attributes. Further refinements are discussed in Sec. 6.2.

3.2 Notable Event Detection

In this section, we explore the detection of notable events from the topic signals. We consider two classes of "notable" events. First, we address detecting changes in processes that are typically stationary: for example, nominal weather that is briefly interrupted by storms. Second, we address detecting anomalies in processes that are non-stationary, but have regular temporal patterns and distributions, such as abnormal traffic congestion caused by holidays or inclement weather.

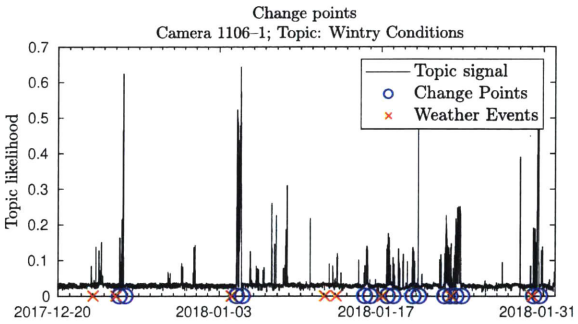
We present statistical techniques and empirical results for addressing each class of events using semantic topic signals. Events of the first class are detected using change-point detection. We demonstrate this in Sec. 3.2.1 by detecting changes in the mean value of the "Wintry Conditions" topic signal to identify inclement weather events. Events of the second type are detected using anomaly detection for samples of data. We demonstrate in Sec. 3.2.2 the detection of anomalous traffic patterns from the "Traffic congestion" topic signal. We validate the performance against known winter storms, holidays, and events.

Furthermore, we examine the performance of using our topic signal representation versus using individual label signals. In particular: we evaluate the performance of the label signals for "blizzard", "rain" and "snow" to serve as benchmarks in the winter storm detection task. For the task of detecting anomalous traffic congestion, we compare to the performance of using the label signals of "traffic", "traffic congestion", "car" and "vehicle."

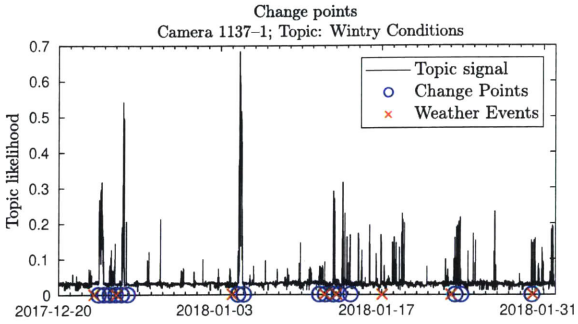
For generality, we will use the notation $\mathbf{X}$ to refer to a set of data. We will consider sets of data constructed from topic signals Θ_c^z (2.4), as well as label signals Λ_c^j (2.1).

3.2.1 Changes in Stationary Processes: Winter Storms

We first consider detecting deviations from stationary processes; that is, processes which typically have a constant mean and variance, but are occasionally disturbed by transitory disruptions. Winter storm disruptions to nominal weather conditions can be modeled as such a process. Under normal weather conditions, a measurement



(a) Camera 1106–1 Topic Signal



(b) Camera 1137–1 Topic Signal

Camera	1106–1			1137–1		
	Prec	Rec	F_1	Prec	Rec	F_1
LS1: snow	0.5	0.5	0.5	0.5	0.5	0.5
LS2: Snow	0.5	0.5	0.5	0.75	0.375	0.5
LS1: blizzard	**1**	0.375	0.5455	**1**	0.25	0.4
LS2: Blizzard	**1**	0.375	0.5455	**1**	0.25	0.4
LS1: rain	**1**	0.375	0.5455	0.857	0.75	0.8
LS2: Rain	**1**	0.375	0.5455	0.857	0.75	0.8
"LS1: rain OR LS1: snow"	0.5	0.5	0.5	0.875	**0.875**	**0.875**
Topic: "Wintry Conditions"	0.625	**0.625**	**0.625**	0.875	**0.875**	**0.875**

(c) Performance evaluation for change-point detection using label signals compared to topics signal. Best scores in each column are rendered in **bold**

Figure 3-2: Performance of winter storm detection using change point detection. Figures in (a)–(b) show the change points in the "Wintry Conditions" topic signal and weather events for cameras 1106–1 and 1137–1. Table (c) compares the performance of using the topic signals to using label signals.

of a "winter storm" process should be constant at zero; however, whenever there is a storm, that process should have a positive, nonzero measurement. If we can construct a signal that replicates this behavior, then we can identify any significant changes in the mean as notable events.

We use change point detection methods to identify the notable events. Change point detection is the problem of finding points in time series where the statistics of the data on either side differ significantly [97]. In our case, we are looking for the points in time where the mean of the preceding and subsequent data differ significantly. This can be posed as an optimization problem [97] of finding the vector ρ of R change

points which minimizes the following objective function:

$$\sum_{r=1}^{R}[\mathcal{C}(\mathbf{X}_{\rho_{r-1}:\rho_r})+B], \tag{3.1}$$

where ρ_r denotes the r^{th} changepoint; $\mathbf{X}_{\rho_{r-1}:\rho_r}$ denotes the data points of dataset $\mathbf{X}$ that fall between the change points ρ_{r-1} and ρ_r; and B is a constant parameter to prevent overfitting. The elements of the change point vector are sequentially ordered in time, i.e. $\rho_r < \rho_{r'}$ iff $r < r'$. The function $\mathcal{C}$ is a cost function which measures the sum of squared differences for some subsample of data, ξ, from its mean μ_ξ:

$$\mathcal{C}(\xi) = \sum_{i=1}^{n}(\xi_i - \mu_\xi)^2$$

where n is the number of elements of ξ.

Essentially, the optimization of (3.1) finds the change points that result in the best fit of a piecewise constant signal to the data. The parameter B prevents overfitting by acting as a minimum threshold: an additional change point is only added if it can reduce the sum of squared difference from the means by at least B. We choose B to be 1/20 of the total energy of the original signal.

We use this technique to find the changes in the "Wintry Conditions" topic signals from cameras 1106–1 and 1137–1, presented in Fig. 3-2a and 3-2b, respectively; i.e. $\mathbf{X} = \Theta_c^z$ for z ="Wintry Conditions" and c =1106–1, and 1137–1. The change points are validated against the eight events in the "Rain" and "Snow" columns in Table 2.1. Since the frequency of the weather data provided by NOAA-GHCN is daily, we allow for a $\pm$12 hour detection window around the change points—i.e. if a weather event happens within a 12 hour window of a detected change point, it is a true positive. This accounts for the temporal uncertainty due to the 24-hour quantization of the reported weather data. Furthermore, we consider pairs of change points as a single detection event: the first change point represents the start of the event (deviation from nominal), and the second represents the end (return to nominal). We also consider additional change points which happen within 24 hours of a start of a detection event

as part of the same event, and thus are not counted as additional change points. This 24-hour minimum duration was chosen to match the NOAA-GHCN data. Finally, we consider (up to) the top eight significant detection events for each signal.

We evaluate the performance of the event detector using the F_1 score, which is the geometric mean between the precision (Prec), and recall (Rec) metrics [24]. Precision measures the fraction of positive classifications which are correct, and recall measures the fraction of total events which are detected. They are given as: $F_1 = \frac{\text{Prec} \cdot \text{Rec}}{\text{Prec}+\text{Rec}}$, where $\text{Prec} = \frac{\text{TP}}{\text{TP}+\text{FP}}$; $\text{Rec} = \frac{\text{TP}}{\text{TP}+\text{FN}}$; TP stands for True Positives; FP for False Positives; and FN for False Negatives.

Fig. 3-2c presents the performance metrics of the changepoint detection applied to the "Wintry Conditions" topic signal and compares it against the use of various label signals. We evaluate the performance of the label signals for "snow," "rain," and "blizzard" from both LS1 and LS2, as well as all pairwise combinations of those labels. We show only the best performing pairwise combination: "LS1: rain OR LS1: snow." We see that in all cases, as measured by F_1 score, the detection events from the topic signal outperform those from the label signals. In addition, it performs as well as, or better than, the performance of the best pairwise combination of labels.

3.2.2 Anomalies in Non-Stationary Processes: Traffic Congestion

Certain processes are inherently non-stationary; for example, the traffic congestion process follows a diurnal pattern of increasing during morning and evening rush hours, and decreasing to zero at night. This non-stationarity prevents us from using the previously discussed change point detection approach to detect notable events. One way to address this would be to model traffic congestion as a trend-stationary process: i.e. a sum of a deterministic time-dependent diurnal trend component and a stationary stochastic component and detect changes in the stochastic component. However, this requires the estimation of the trend component, which introduces another modeling and statistical question.

Instead, we present an alternative approach which sidesteps the need to estimate the temporal trend signal. We pose the problem as a statistical anomaly detection problem by measuring the dissimilarity (via an f-divergence measure) between the empirical distribution of the signal values and a set of nominal reference distributions. Furthermore, we employ a direct estimation technique [34, 18] for computing the divergence between two empirical distributions without having to parametrically estimate the distributions themselves as well. Our method offers significant generality, as it does not depend on the functional forms of the temporal trend or distribution.

In addition, we consider data *subsequences* as our "data points." A subsequence starting at data point $x_i \in \mathbf{X}$ is represented as $\chi_i = [x_i, x_{i+1}, \ldots, x_{i+k-1}] \in \mathbb{R}^{Mk}$, where k is the subsequence length and M is the number of dimensions of x_i. We then use the set of subsequences $\boldsymbol{\chi} = \{\chi_i\}_{i=1}^{N_{\text{img}}-k}$ as the dataset for anomaly detection. We note that this process is similar to the construction of the lag terms in autoregressive (AR) models, and is used in other time series analysis problems [37]. We vary the length subsequence length k and empirically determine the best subsequence length to use in our anomaly detection procedure.

Our approach considers the anomaly detection problem of whether a test sample of data from a signal is *anomalous* compared to a set of known-nominal reference samples. Let us divide the length of a signal into equal-sized time windows, where T_s denotes s^{th} window and $s \in [1, N_{\text{win}}]$. Let $\mathbf{X}_s$ denote a *test* sample of data corresponding to the data points which occur during T_s. Let $\overline{\mathbf{X}}_\sigma$ denote a nominal reference sample, where $\sigma \in [1, N'_{\text{win}}]$ indexes the reference samples, and the set of all reference samples is denoted $\overline{\mathbf{X}}$. We determine whether a sample $\mathbf{X}_s$ is anomalous based on its average dissimilarity to the reference samples $\overline{\mathbf{X}}_\sigma \in \overline{\mathbf{X}}$.

Divergence Measures and Estimation

We compute dissimilarity between two data distributions using an f-divergence, defined as follows: for probability distributions P, P', defined over a space Ω (with

respective probability densities $p(\mathbf{x}), p'(\mathbf{x})$), an f-divergence from P to Q is given by:

$$D_{\mathrm{f}}(P||P') := \int_{\Omega} p'(\mathbf{x}) f\left(\frac{p(\mathbf{x})}{p'(\mathbf{x})}\right) d\mathbf{x} \tag{3.2}$$

where $f(t)$ is a convex function with $f(1) = 0$ [98]. The well-known Kullback-Leibler (KL) divergence [99], and Pearson (PE) χ^2-divergence [100] are specific instances of f-divergences, where $f_{KL}(t) = t\log(t)$ and $f_{PE}(t) = \frac{1}{2}(t-1)^2$ respectively.

All f-divergences are positive, are minimized at zero when P and P' are identical, and maximized when they are statistically independent; in addition, they satisfy information monotonicity and joint convexity [98]. A larger f-divergence value indicates a greater dissimilarity between two distributions than a smaller divergence; note however, that f-divergences are not true distance measures, in that they are not commutative, i.e. $D_f(P||P') \neq D_f(P'||P)$, and do not satisfy the triangle inequality. In our application, we adopt the common practice [37] of using a symmetrized divergence which satisfies commutativity, given as $D_{\mathrm{f}}^{\mathrm{sym}}(P||P') := D_{\mathrm{f}}(P||P') + D_{\mathrm{f}}(P'||P)$.

In this dissertation, we consider a variant of the PE divergence, the Relative Pearson (RP) divergence [18], defined as:

$$D_{\mathrm{RP}}(P||P') = \frac{1}{2}\int_{\Omega} q_{\gamma}(\mathbf{x})\left(\frac{p(\mathbf{x})}{q_{\gamma}(\mathbf{x})} - 1\right)^2 d\mathbf{x} \tag{3.3}$$

where

$$q_{\gamma}(\mathbf{x}) = \gamma p(\mathbf{x}) + (1-\gamma)p'(\mathbf{x}), \tag{3.4}$$

for some $\gamma \in [0,1)$, is referred to as the γ-relative density [18][2]. The use of $q_{\gamma}(\mathbf{x})$ in (3.3) ensures that the *γ-relative density ratio*, $r_{\gamma}(\mathbf{x}) = p(\mathbf{x})/q_{\gamma}(\mathbf{x})$, stays upper bounded by $\frac{1}{\gamma}$. This boundedness improves the rate of numerical convergence when estimating the divergence [18].

We estimate the divergence using the Relative unconstrained Least-Squares Im-

[2]The notation in [18] refers to this quantity as the α-relative density. However, in this dissertation, we refer to it as the γ-relative density to avoid the ambiguity with the α LDA hyperparameter.

portance Fitting (RuLSIF) direct estimation procedure presented in [18]. RuLSIF is an extension of direct divergence-estimation procedures such as the KLIEP for estimating KL divergence [34] and unconstrained Least-Squares Importance Fitting (uLSIF) for estimating the Pearson divergence [35]. These direct estimation procedures estimate the divergence measure between two data samples directly from the data without the need to parametrically estimate the two distributions $p(\mathbf{x})$ and $p'(\mathbf{x})$. This is quicker to compute and more accurate in estimating divergences than estimating the distributions separately [34, 35, 18]. We choose RuLSIF in particular because it has a better convergence rate compared to uLSIF and KLIEP [18].

The RuLSIF procedure for estimating the RP divergence between two distributions—P, P'—given empirical samples from each [18], is given below.

Consider two sets of independent and identically distributed (i.i.d.) samples X, X' drawn from P and P', denoted as:

$$X = \{\mathbf{x}_i\}_{i=1}^{n} \overset{i.i.d.}{\sim} P,$$
$$X' = \{\mathbf{x}'_j\}_{j=1}^{n'} \overset{i.i.d.}{\sim} P',$$

where n, and n' are the respective number of draws in each sample.

The Relative Pearson divergence (3.3) can be written in terms of the γ-relative density ratio, $r_\gamma(\mathbf{x})$:

$$\begin{aligned} D_{\mathrm{RP}}(P||P') &= \frac{1}{2}\mathbb{E}_{q_\gamma(\mathbf{x})}[(r_\gamma(\mathbf{x})-1)^2] \\ &= \frac{1}{2}\left[\int (r_\gamma(\mathbf{x})^2 - 2r_\gamma(\mathbf{x}) + 1) q_\gamma(\mathbf{x}) d\mathbf{x}\right] \\ &= \frac{1}{2}\left[\int r_\gamma(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} - 2\int (r_\gamma + 1) q_\gamma(\mathbf{x}) d\mathbf{x}\right] \\ &= \frac{1}{2}\mathbb{E}_{p(\mathbf{x})}[r_\gamma(\mathbf{x})] - \frac{1}{2} \end{aligned}$$

where $\mathbb{E}_{p(\mathbf{x})}[f(\mathbf{x})]$ denotes the expectation of $f(\mathbf{x})$ over $p(\mathbf{x})$. Using the empirical

sample average to estimate the expectation yields:

$$D_{\mathrm{RP}}(P||P') \approx \frac{1}{2n}\sum_{i=1}^{n} r_\gamma(\mathbf{x}_i) - \frac{1}{2}. \tag{3.5}$$

The γ-relative density ratio, $r_\gamma(\mathbf{x})$, is estimated with a kernel model, $g(\mathbf{x};\xi)$, with Gaussian kernel basis functions, $K(\mathbf{x},\mathbf{x}_l)$, given as:

$$g(\mathbf{x};\xi) := \sum_{l=1}^{n} \xi_l K(\mathbf{x},\mathbf{x}_l),$$
$$K(\mathbf{x},\mathbf{x}') = \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\sigma^2}\right),$$

where ξ is an n-dimensional vector of parameters to be inferred from the data, $^\top$ denotes the transpose operator, and $\sigma > 0$ is the kernel width parameter.

The ξ parameter vector is chosen to minimize the expected squared error $J(\xi)$ between $g(\mathbf{x};\xi), r_\gamma(\mathbf{x})$:

$$\begin{aligned} J(\xi) &:= \frac{1}{2}\mathbb{E}_{q_\gamma(\mathbf{x})}\left[(g(\mathbf{x};\xi) - r_\gamma(\mathbf{x}))^2\right] \\ &= \frac{\gamma}{2}\mathbb{E}_{p(\mathbf{x})}\left[g(\mathbf{x};\xi)^2\right] + \frac{1-\gamma}{2}\mathbb{E}_{p(\mathbf{x})}\left[g(\mathbf{x};\xi)^2\right] \\ &\quad + \mathbb{E}_{p(\mathbf{x})}\left[g(\mathbf{x};\xi)\right] + \text{Constant}. \end{aligned}$$

This problem is posed as an L_2-regularized least squares problem of the form:

$$\widehat{\xi} := \underset{\xi\in\mathbb{R}^n}{\arg\min}\left[\frac{1}{2}\xi^\top \widehat{H}\xi - \widehat{h}^\top\xi + \frac{\eta}{2}\xi^\top\xi\right] \tag{3.6}$$

where $\widehat{H}$ is an $n\times n$ matrix with the element (l,l') given by (3.7), $\widehat{h}$ is an n-dimensional

vector with element l given by (3.8), and $\eta \geq 0$ is a regularization parameter:

$$\widehat{H}_{l,l'} := \frac{\gamma}{n} \sum_{i=1}^{n} K(\mathbf{x}_i, \mathbf{x}_l) K(\mathbf{x}_i, \mathbf{x}_{l'}) + \frac{1-\gamma}{n'} \sum_{j=1}^{n'} K(\mathbf{x}'_j, \mathbf{x}_l) K(\mathbf{x}'_j, \mathbf{x}_{l'}) \tag{3.7}$$

$$\widehat{h}_l := \frac{1}{n} \sum_{i=1}^{n} K(\mathbf{x}_i, \mathbf{x}_l) \tag{3.8}$$

The L_2-regularized least squares problem in (3.6) admits an analytic solution of the form:

$$\widehat{\xi} = (\widehat{H} + \eta \mathbf{I}_n)^{-1} \widehat{h},$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix. Thus, the estimator for the γ-relative density ratio $r_\gamma(x)$ is given by (3.9). Plugging (3.9) into (3.5) yields our estimator, $\widehat{D_{\mathrm{RP}}}(\mathbf{X}, \mathbf{X}')$, for the Relative Pearson divergence (3.10):

$$\widehat{r}_\gamma(\mathbf{x}) := g(\mathbf{x}; \widehat{\xi}) = \sum_{l=1}^{n} \widehat{\xi}_l K(\mathbf{x}, \mathbf{x}_l) \tag{3.9}$$

$$\widehat{D_{\mathrm{RP}}}(\mathbf{X}, \mathbf{X}') := \frac{1}{2n} \sum_{i=1}^{n} \widehat{r}_\gamma(\mathbf{x}_i) - \frac{1}{2}. \tag{3.10}$$

The kernel width, σ, and regularization parameter, η, are selected as those that minimize the leave-one-out cross-validation error.

Relative Pearson Divergence Anomaly Score

We construct an anomaly score for a test sample $\mathbf{X}_s$ and set of reference samples $\overline{\mathbf{X}}$ based on the Relative Pearson divergence. We refer to this anomaly score as the Relative Pearson Divergence Anomaly Score (RPDAS). It is computed as the average symmetrized RP divergence between the test sample and each of the reference

samples, given as:

$$\text{RPDAS}(\mathbf{X}_s, \overline{\mathbf{X}}) := \frac{1}{N'_{\text{win}}} \sum_{\sigma=1}^{N'_{\text{win}}} \widehat{D}_{\text{RP}}^{\text{sym}}(\mathbf{X}_s, \overline{\mathbf{X}}_\sigma), \tag{3.11}$$

where $\widehat{D}_{\text{RP}}^{\text{sym}}(\mathbf{X}_s, \overline{\mathbf{X}}_\sigma)$ is the symmetrized, RuLSIF-estimated RP divergence between the test sample $\mathbf{X}_s$ and reference sample $\overline{\mathbf{X}}_\sigma$. The RPDAS is bounded on the same range as the RP divergence: $[0, 1/\gamma]$.

A sample $\mathbf{X}_s$ is flagged as a *detection event* if the RPDAS of that sample exceeds an alert threshold τ. An anomaly detection event for sample $\mathbf{X}_s$ is considered a true positive detection if there is a true *anomaly event* during the time period T_s. False positives correspond to detection events without a corresponding true anomaly event, and false negatives correspond to true anomaly events without a detection event. By varying the threshold $\tau \in [0, 1/\gamma]$, we can adjust the sensitivity of the anomaly detection process. In this way, we compute a Precision-Recall (PR) curve [24] to evaluate performance. We use both the area under the PR curve (PR AUC), as well as the configuration with the best F_1 score as performance metrics. The PR AUC evaluates the overall performance of the anomaly detector (with 1 being the best possible PR AUC score), while the best F_1 score indicates the best-case performance corresponding to an optimally-calibrated detector [24].

Empirical Validation

We now validate our approach to detecting notable events via anomaly detection on topic signals. We consider the process of traffic congestion, as measured by the "Traffic congestion" topic signal. We focus on the data from camera 1508–1, which was the only camera in the dataset with no changes in camera angle or perspective. In addition, its location in an underpass protects it from atmospheric interference and obstruction. These properties help ensure that any variations in traffic congestion that we detect reflect the actual changes in in the process, and not caused by outside factors.

We detect days with anomalous distributions of data points and subsequences

in the topic signal $\Theta_{1508\text{–}1}^{\text{Traffic congestion}}$ using the RPDAS, and compare these detection events with known anomaly events which we expect to significantly affect traffic congestion. These anomaly events are the "holiday/special event" and "snow" columns of Table 2.1. We do not consider the rain-only events, as we believe they were less likely to cause disruptions to traffic compared to snowfall, holidays, or special events. In the validation analysis, we do not observe that the detector process is sensitive to rain-only events. To match the daily temporal granularity of the anomaly event data in Table 2.1, the test samples are partitioned as 24-hour-long windows.

Data from phase I is used as the reference dataset $\overline{\mathbf{X}}$. The reference data spans November 6^{th}–November 12^{th}, 2017, contained no significant weather events or holidays, and was also partitioned as 24-hour-long windows.

The daily RPDAS is computed for $\Theta_{1508\text{–}1}^{\text{Traffic congestion}}$, with $\gamma = 10^{-3}$ for the γ-relative density parameter, and for various subsequence window lengths $k \in \{1, 2, 4, 8\}$. The threshold τ was varied from zero to $1/\gamma$ to construct Precision-Recall curves. The PR curves for all configurations of (k, τ) are presented in Fig. 3-3a. The baseline performance of a null classifier (i.e. guessing uniformly at random) is indicated with the red horizontal line, which has a corresponding PR AUC of 0.14.

Fig. 3-4 shows the daily RPDAS for the configuration with the best F_1 score ($k = 2$, $\tau^* = 21$). We see that it performs reasonably well: it has one missed detection of New Year's Day, and two false positives: one the weekend before Christmas, and one right after the bomb cyclone storm.

For comparison, we also computed the PR curves for a number of individual label signals which we expected might capture the traffic congestion process, including "car," "vehicle," "traffic," and "traffic congestion" from both LS1 and LS2. Figures 3-3b–3-3e displays the PR curves for each label (for brevity, we only show the better performing label between LS1 and LS2). We found that no individual label signal achieved comparable performance in anomaly detection in PR AUC or best F_1 score. Furthermore, in cases such as in Fig. 3-3c and 3-3c, anomaly detection on the label signals performs worse than the null classifier.

The fact that the topic signal outperforms any individual label signal in the val-

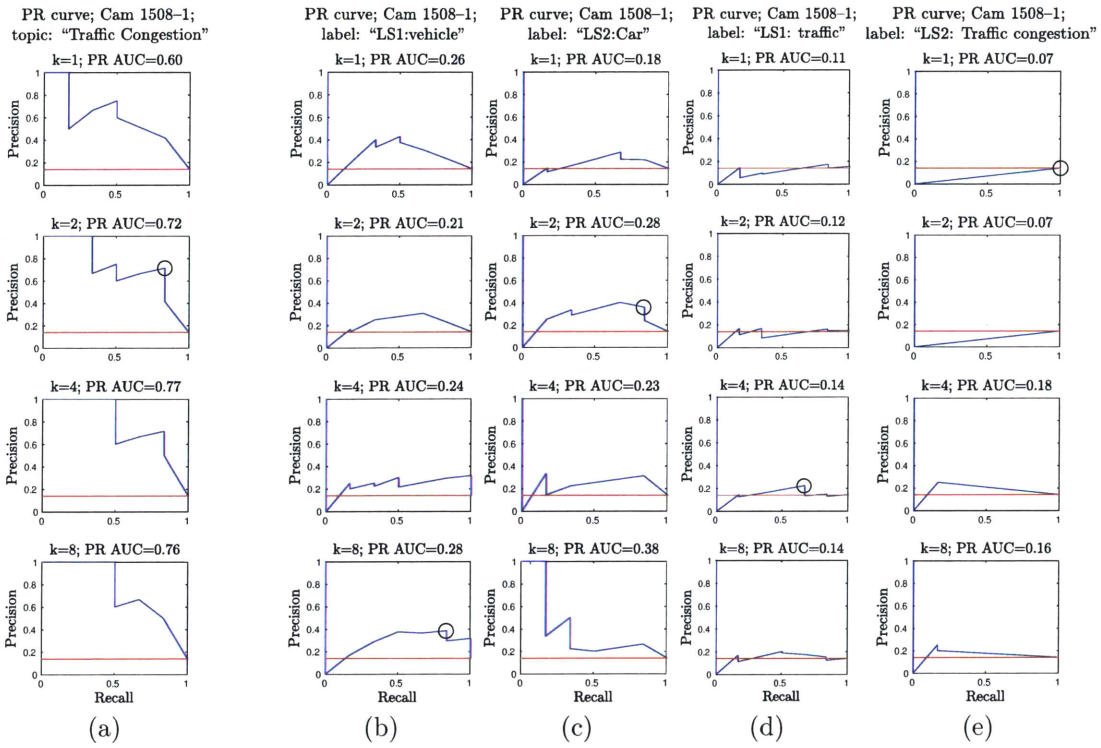


Figure 3-3: Precision-recall curves for anomalous traffic detection for various signals and subsequence window lengths k. The black circle indicates highest F_1 score in each column; the horizontal red line indicates performance of null predictor. Fig. (a) shows the results using the LDA "Traffic congestion" topic signal, whereas Figs. (b)–(e) show the results using a number of selected label signals.

idation tasks demonstrates that the performance of the topic signal is not simply due to the performance of its component label signals. Instead, topics capture additional information in *combinations and distributions* of labels that isn't captured by individual labels.

3.3 Part I Discussion

In this part, we introduced the BFCC dataset, labeled with semantic feature labels; the BoLW semantic feature vector model; the use of the LDA topic model for extracting semantic topic signals; applications of the BoLW representation for image retrieval; and signal processing-style analyses of topic and label signals for notable event detection.

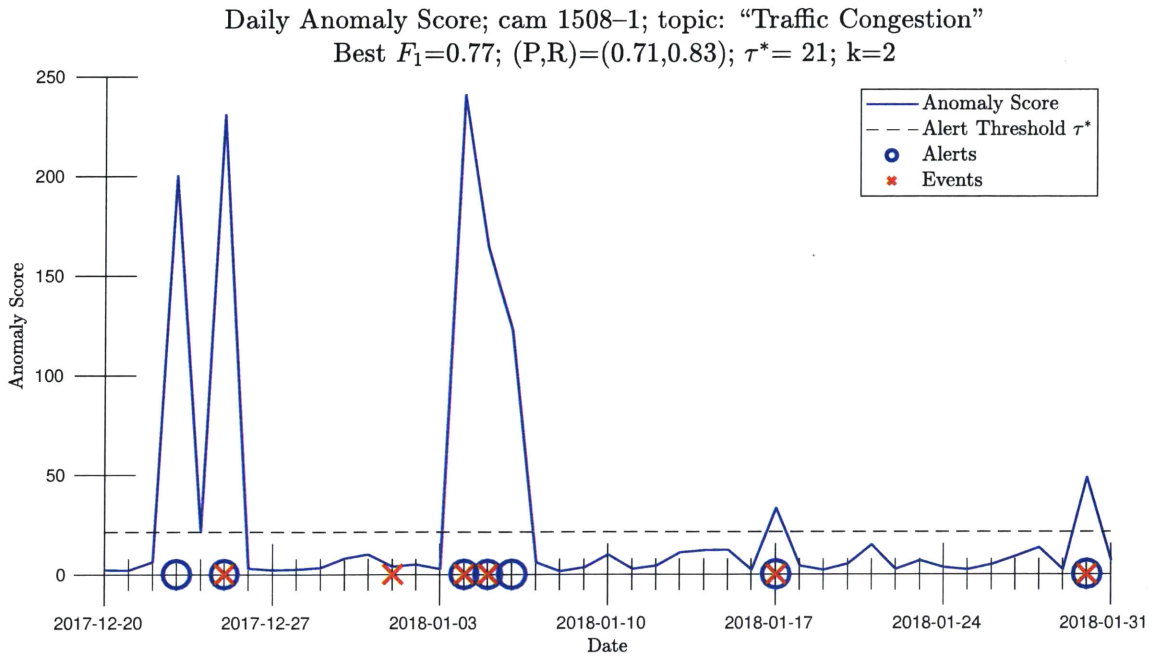


Figure 3-4: Anomalous traffic detection results for Camera 1508–1. The figure renders the daily anomaly score as a blue line; the alert threshold τ^* is depicted as the horizontal dashed line; anomaly events are denoted with a red "x", and alerts are denoted with a blue "o".

The main contributions of this part are in demonstrating the utility of using semantic features from common implementations of image labeling software for machine learning and signal processing applications. The semantic representations provide a significantly reduced-dimension representation of complex visual elements and phenomena, such as traffic and weather, as a conventional time series. This allows for footage to be reviewed at a much larger scale than what its currently possible with human observation.

Notably, the semantic representations are significantly compressed compared to the original images. This can be beneficial for longitudinal analyses, since the semantic representation requires much less space to store and less computation to analyze. Our work demonstrates that even using only semantic information, we can identify the onset of traffic and weather disruptions.

Furthermore, semantic representations provides additional privacy-related benefits, since the semantic representations are less likely to capture personally identifying information. On the other hand, this compression and lossiness of information may be

a trade-off in applications that require precision, such as in robotics and autonomous vehicles. Thus, we believe that semantic representations are not a replacement for other representations and sensors, but instead a complement. Indeed, we believe that there is a promising avenue of future work in fusing semantic features with other sensing modalities, such as visual and multispectral imagery, as well as other traffic infrastructure sensors such as loop detectors and radar.

Part II

Bayesian Congestion Games

Part II of this dissertation considers game theoretic models of routing decisions where travelers have heterogeneous information about the state of a disruption-prone network. Game theoretic models, particularly congestion games, provide an analytical framework for predicting the behavior of travelers in a congested network, where the cost of traversing a link increases with the number of travelers using that link. The availability of information about traffic conditions creates a feedback loop which affects the decisions of travelers, which should be considered when modeling route choices. Figure 3-5 illustrates how the content of Part II relates to information feedback loop. In particular, we examine the effects of information access, accuracy, and *subjective* perception of others' information on the equilibrium route choices of travelers.

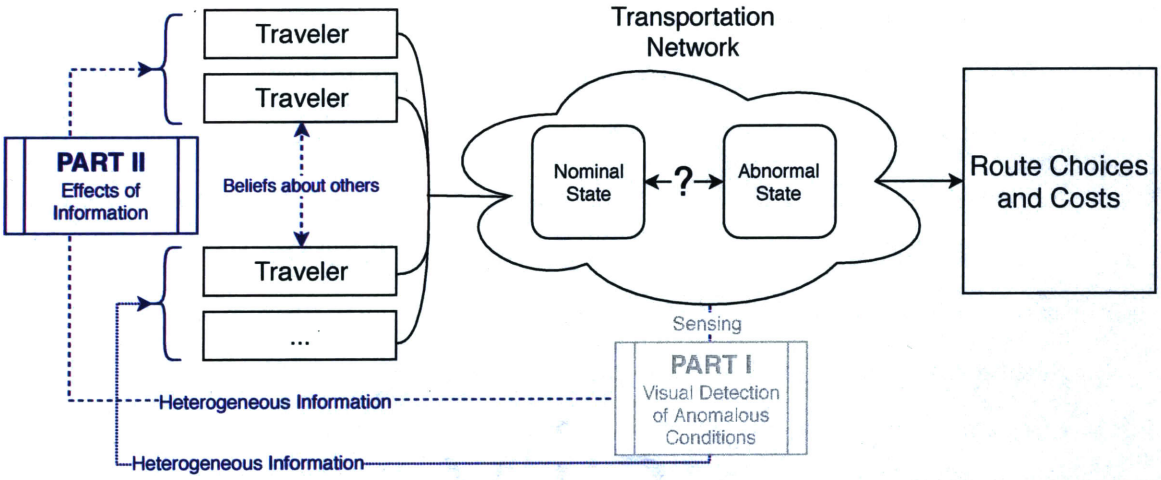


Figure 3-5: Part II addresses game theoretic models of traveler route decisions with incomplete information and potentially subjective beliefs.

The need to consider information as well as the *perception* of others' information is illustrated with the following thought experiment. Consider the decision between taking two similar-length routes to a destination. If a traveler has no information about the state of each route, then they may be indifferent between either route. But if the traveler learns that there is an accident on the first route, then they may be more inclined to take the second route. Yet, the decision to take the second route also depends on what information they believe other travelers have received. If they believe that they are the only traveler with the information about the accident, then

they are likely to take the second route. On the other hand, if they believe that every other traveler also knows that information, then they are less likely to take the second route, due to the additional congestion that would be caused by the other travelers who choose to switch to the second route. Thus, both the information about the accident, as well as the perception of others' information about the accident need to be taken into account when modeling the route choices of all travelers.

While travelers could have many possible beliefs about the information of other travelers, there is only one "true" information distribution. The conventional Bayesian Game formulations assume that all players' beliefs are consistent with this true information distribution (the common prior assumption). Yet, it is not a stretch to imagine that travelers may not accurately know about the information that other travelers have. Thus, we present a formulation and solutions of *Subjective* Bayesian Congestion Games which relaxes the common prior assumption to allow for a broader set of traveler beliefs.

Traditional, complete information congestion games have been thoroughly studied [101, 42, 52, 40], and have been used to provide bounds on the inefficiency of user equilibrium versus the social optimum route flows (the so-called "price of anarchy") [53]. However, the conventional formulations do not account for stochasticity in the state of the network, such as the presence and location of traffic incidents. Furthermore, different travelers may have different information about the network state, as well as about other travelers. We present Bayesian extensions to the congestion game framework which accounts for stochastic cost functions as well as heterogeneous, incomplete information about the state of the network and other travelers' information.

Chapter 4 presents the formulations of the Objective and Subjective Bayesian Congestion Game. The Objective BCG adopts the conventional common prior assumption. The common prior assumption requires that the beliefs of travelers be consistent with the true probabilities used by "Nature." The Objective BCG is a natural extension of congestion games to an incomplete information context. In addition, the adoption of the common prior allows it to be easily solved as the solution of a convex optimization problem. This provides a theoretical basis and starting point

for investigating the effects of information distribution and accuracy (with regards to the state of the network) on the equilibrium route choices of travelers.

The Subjective BCG presented in this dissertation accommodates beliefs that are not derived from the common prior—allowing for the modeling of a broader range of beliefs with regards to other travelers' information. Our formulation assumes a bounded rationality view of subjective beliefs: that travelers have a misperceived belief of the information available to other travelers, but behave rationally conditioned on that belief. This is distinct from the approach of SUE models, which capture all irrationality as a random variable of unexplained preferences.

Chapter 5 presents the equilibrium results of the Objective and Subjective BCGs. The Bayesian Wardrop Equilibrium is used as the solution concept for both formulations. We provide the equilibrium route flows, and characterize the effects of subjectivity on the equilibrium route choices and costs. We show that there exists three distinct classes of Subjective BCG equilibrium, which are neatly characterized by a "misperception factor." One of these equilibrium classes is shown to be equivalent to the solution of the conventional Objective BCG. We examine the effect of each of the parameters—incident probability, the fraction of travelers with information, and perceived accuracy of information—on each individual population's costs, as well as the average cost of all travelers. We note that the subjective perception of other travelers' information has a significant effect resulting in qualitatively distinct equilibrium structure, as well as distinctly different individual and social costs in equilibrium compared to the conventional objective case.

Whereas the Objective BCG can be solved as a convex optimization problem, the Subjective BCG cannot be solved in the same way. However, the Bayesian Wardrop Equilibrium conditions can also be written as a (potentially infinite) set of variational inequalities. Yet, we demonstrate that the equilibrium conditions for the Subjective BCG can be written as a finite number of variational inequalities. The equilibrium can then be found as the solution of a non-convex optimization problem, using the Objective BCG as an initial guess.

This work is based off the published and submitted papers in [102, 103]. Comple-

mentary work focusing on the Objective BCG was also done by collaborators Wu et al. in [104]. We also note that the notation in this part is independent of the notation from Part I.

Chapter 4

Bayesian Congestion Game Formulations

In this chapter, we formulate two flavors of the BCG: Objective and Subjective. We present in Sec. 4.3 the Objective BCG: a Bayesian extension of the complete information congestion game into an incomplete information context and adopts the conventional common prior assumption. This game is called "objective" because the beliefs of the travelers are consistent with the "true" probability distribution used by the fictitious player Nature to instantiate the game and signals. The Subjective BCG, presented in Sec. 4.4, is a more general formulation which relaxes the common prior assumption and allows for beliefs that are not consistent with the distribution used by Nature.

Both formulations utilize the same environment, presented in Sec. 4.1. Sec. 4.2 provides the information structure and beliefs of the travelers; and Sec. 4.3 provides the formal formulation, as well as solution concept. The analysis of equilibrium is presented jointly with the analysis of the Subjective BCG in the following chapter.

4.1 Incident-Prone Network

Consider a traffic network with a single origin-destination pair connected by 2 parallel routes. Define the set of routes as $R = \{\mathrm{r}_1, \mathrm{r}_2\}$. There is a fixed demand D of

nonatomic travelers who make route-choice decisions on the network. Travelers incur a cost that depends on the aggregate load on their chosen route, as well as the "state of the network". The network state is a random variable, s, which models random disruptions, such as traffic accidents, that may occur on the network.

In this dissertation, we restrict our attention to a particular configuration of network states: we assume that only route r_1 is prone to random incidents, and if there is an incident on route r_1, the travel time cost of r_1 is higher than that in the normal (i.e. non-incident) state with the same aggregate load of travelers on that route. The cost function of route r_2 is not affected by the state. The random state of the network, denoted s, is defined as:

$$s \triangleq \begin{cases} \mathrm{a} & \text{if there is an incident on } \mathrm{r}_1 \text{ (abnormal state)} \\ \mathrm{n} & \text{if there is no incident on } \mathrm{r}_1 \text{ (normal state).} \end{cases} \tag{4.1}$$

The set of states is $S \triangleq \{\mathrm{a}, \mathrm{n}\}$. The realization of the state s is drawn by a fictitious player—"Nature"—according to an exogenous and fixed prior distribution, denoted θ. The state distribution θ follows a Bernoulli distribution, with probability of an incident on route r_1 $\theta(\mathrm{a}) = p$, and complementarily, the probability of no incident $\theta(\mathrm{n}) = 1 - p$. Once the state is realized, it is fixed throughout the game.

We consider route cost functions where the travel time is an affine function of the load. The slope of route r_1 is *state-dependent*; we denote the slope of r_1 in state a (resp. n) as α_1^{a} (resp. α_1^{n}). The slope of route r_2, denoted α_2, as well as the free flow travel time of each route, denoted β_{r}, is not dependent on the state. The route cost functions are thus given by:

$$\begin{aligned} c_1^s(l_1) &= \begin{cases} \alpha_1^{\mathrm{a}} \cdot l_1 + \beta_1 & \text{if } s = \mathrm{a} \\ \alpha_1^{\mathrm{n}} \cdot l_1 + \beta_1 & \text{if } s = \mathrm{n}, \end{cases} \\ c_2(l_2) &= \alpha_2 \cdot l_2 + \beta_2. \end{aligned} \tag{4.2}$$

where l_{r} is the aggregate load on route r.

Additionally, we assume that the congestion level of route r_1 is higher than that

of r_2 when there is an accident, but lower when the state is normal, i.e. $\alpha_1^a > \alpha_2 > \alpha_1^n$. The free flow travel time of r_1 is less than that of r_2, i.e. $\beta_1 < \beta_2$. The network and the cost functions are illustrated in Figure 4-1.

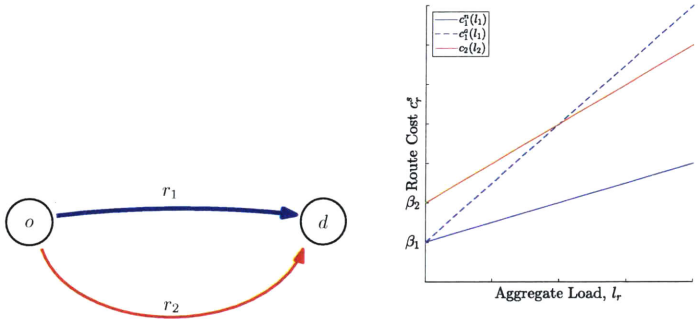


Figure 4-1: Network diagram and route cost functions

4.2 Information and Beliefs

Travelers are divided into two populations: population H ("Higher accuracy") and population L ("Lower accuracy"). The fraction of travelers in population H and L are respectively denoted as λ^H and λ^L, where $\lambda^L = 1 - \lambda^H$.

Each population $i \in I \triangleq \{H, L\}$ receives private information of the state, denoted $t^i \in T^i \triangleq \{a, n\}$. Additionally, we assume that t^H and t^L are independent with each other, conditional on state s:

$$\Pr(t^H, t^L | s) = \Pr(t^H | s) \cdot \Pr(t^L | s), \quad \forall t^H \in T^H, \quad \forall t^L \in T^L, \quad \forall s \in S. \tag{4.3}$$

The accuracies of each player's type signals are given by the parameters δ^i, respectively

$$\Pr(t^H = s | s) = \delta^H, \quad \Pr(t^H \neq s | s) = 1 - \delta^H, \quad \forall s \in S, \tag{4.4a}$$

$$\Pr(t^L = s | s) = \delta^L, \quad \Pr(t^L \neq s | s) = 1 - \delta^L, \quad \forall s \in S. \tag{4.4b}$$

Importantly, each population i knows the prior distribution of the state θ, the accuracy of their own information $\Pr(t^i | s)$ as in (4.4), and may or may not know

the accuracy of the other population's information, $\Pr(t^{-i}|s)$. In the Objective BCG, formulated in Sec. 4.3 the common prior assumption assumes that $\Pr(t^{-i}|s)$ is public knowledge. In Sec. 4.4, we provide the formulation for the case where $\Pr(t^{-i}|s)$ is not public knowledge, and instead each population has their own belief $\Pr^i(t^{-i}|s)$.

Based on the knowledge of the prior, one's own information accuracy, and the accuracy of the other population's information, each population forms a belief about the state s and the opponent's information t^{-i} given their own information t^i, denoted as $\mu^i(s,t^{-i}|t^i)$, which can be written as follows:

$$\mu^i(s,t^{-i}|t^i) = \frac{\Pr^i(s,t^i,t^{-i})}{\Pr(t^i)} = \frac{\theta(s)\Pr^i(t^i,t^{-i}|s)}{\Pr(t^i)}, \quad \forall s \in s, \quad \forall t^i \in T^i, \quad \forall t^{-i} \in T^{-i}, \quad \forall i \in I, \tag{4.5}$$

where $\Pr(t^i) = \sum_{s\in S}\theta(s)\Pr(t^i|s)$ is the marginal probability of t^i. Since t^i and t^{-i} are independent, conditioned on the state s, and each population knows their own information accuracy, we can simplify (4.5) to:

$$\mu^i(s,t^{-i}|t^i) = \frac{\theta(s)\Pr(t^i|s)}{\Pr(t^i)} \cdot \Pr(t^{-i}|s), \quad \forall s \in s, \quad \forall t^i \in T^i, \quad \forall t^{-i} \in T^{-i}, \quad \forall i \in I \tag{4.6}$$

in the Objective BCG, and

$$\widetilde{\mu}^i(s,t^{-i}|t^i) = \frac{\theta(s)\Pr(t^i|s)}{\Pr(t^i)} \cdot \Pr^i(t^{-i}|s), \quad \forall s \in s, \quad \forall t^i \in T^i, \quad \forall t^{-i} \in T^{-i}, \quad \forall i \in I \tag{4.7}$$

for the Subjective BCG. From (4.7), we can see that the belief $\mu^i(s,t^{-i}|t^i)$ is comprised of two parts: (1) The posterior distribution of the state $\Pr(s|t^i)$, which depends on their own information accuracy; (2) The belief of the other population's conditional type distribution: $\Pr(t^{-i}|s)$ in the objective case, and $\Pr^i(t^{-i}|s)$ in the subjective. Note that the Objective BCG is an edge case of Subjective BCG, where $\Pr^i(t^{-i}|s) = \Pr(t^{-i}|s)$.

4.3 Game Formulation and Equilibrium Concept

In the framework of Bayesian games (see [73] and [50]), the notion of "type" encodes the private information available to each population. In our model, the type of population i is the private information t^i. The type set is T^i. The belief of the state $s \in S$ and the opponent's type $t^{-i} \in T^{-i}$ is $\mu^i(s, t^{-i}|t^i)$ as in (4.7).

A strategy of population $i \in I$ is an $|R|$-dimensional vector function of their type, denoted as $q^i(t^i) = (q^i_{\mathrm{r}}(t^i))_{\mathrm{r}\in R, t^i \in T^i}$, where $q^i_{\mathrm{r}}(t^i)$ is the demand assigned by population i to the route r when their type is t^i. We say that the strategy q^i is feasible if and only if it satisfies:

$$\sum_{\mathrm{r}\in R} q^i_{\mathrm{r}}(t^i) = \lambda^i D, \quad \forall t^i \in T^i, \quad \forall i \in \{\mathrm{H}, \mathrm{L}\}, \tag{4.8a}$$

$$q^i_{\mathrm{r}}(t^i) \geq 0, \quad \forall \mathrm{r} \in R, \quad \forall t^i \in T^i, \quad \forall i \in \{\mathrm{H}, \mathrm{L}\}, \tag{4.8b}$$

which correspond to the requirements that all demand be routed on the network (4.8a), and that the demand on each route be non-negative (4.8b).

We denote the set of feasible strategies for population i as Q^i, and denote a feasible strategy profile of all players by $q \triangleq (q^{\mathrm{H}}, q^{\mathrm{L}}) \in Q^{\mathrm{H}} \times Q^{\mathrm{L}}$. The aggregate load on route r that is induced by strategy profile $(q^{\mathrm{H}}, q^{\mathrm{L}})$ is denoted as:

$$l_{\mathrm{r}}(t^{\mathrm{H}}, t^{\mathrm{L}}) = q^{\mathrm{H}}_{\mathrm{r}}(t^{\mathrm{H}}) + q^{\mathrm{L}}_{\mathrm{r}}(t^{\mathrm{L}}), \quad \forall \mathrm{r} \in R, \quad \forall t^{\mathrm{H}} \in T^{\mathrm{H}}, \quad \forall t^{\mathrm{L}} \in T^{\mathrm{L}}. \tag{4.9}$$

We are now ready to formally define the Bayesian congestion game Γ:

$$\Gamma \triangleq (I, S, T, Q, C, \mu), \tag{4.10}$$

where:

$I = \{\mathrm{H}, \mathrm{L}\}$: Set of populations with generic member $i \in I$

$S = \{\mathrm{a}, \mathrm{n}\}$: Finite set of network states with element $s \in S$

$T = (T^i)_{i\in I}$: Set of player types, with population i's type $t^i \in T^i$

$Q = (Q^i)_{i \in I}$: Set of feasible strategy profiles, $q = (q^{\mathrm{H}}, q^{\mathrm{L}}) \in Q$

$C = \{c_{\mathrm{r}}^s(\cdot)\}_{\mathrm{r} \in R, s \in S}$: Set of state-dependent route cost functions, with c_{r}^s as defined in (4.2)

$\mu = \left(\mu^i\right)_{i \in I}$: Set of beliefs, where $\mu^i = \left(\mu^i(s, t^{-i} | t^i)\right)_{s \in S, t^{-i} \in T^{-i}, t^i \in T^i}$ as defined in (4.7);

The common (public) knowledge of the game includes: the total demand D; the parameter λ^{H} describing the relative population sizes; the set of routes R and the corresponding route cost functions $\{c_{\mathrm{r}}^s\}_{s \in S, \mathrm{r}, \in R}$; the set of network states S and its prior distribution $\theta(s)$, and the type spaces $\left(T^i\right)_{i \in I}$. In addition, for the objective formulation in this chapter, players also know the opponent populations' conditional type distribution $\Pr(t^{-i}|s)$. In the Subjective BCG formulation of the next chapter, this is not common knowledge.

Given a feasible strategy profile, $q \in Q$, the expected cost of route r based on the belief of population i given their type t^i can be written as

$$\mathbb{E}_\mu[c_{\mathrm{r}}(q)|t^i] = \sum_{s \in S} \sum_{t^{-i} \in T^{-i}} \mu^i(s, t^{-i}|t^i) c_{\mathrm{r}}^s(q_{\mathrm{r}}^i(t^i) + q_{\mathrm{r}}^{-i}(t^{-i})), \quad \forall \mathrm{r} \in R, \quad t^i \in T^i, \quad i \in I. \tag{4.11}$$

We use the Bayesian Wardrop Equilibrium as the solution concept, which is defined as follows:

Definition 1. *Bayesian Wardrop Equilibrium (BWE)*
The feasible strategy profile q^* is a (BWE), if for any $i \in I$ and any $t^i \in T^i$:

$$q_{\mathrm{r}}^{i*}(t^i) > 0, \quad \Rightarrow \quad \mathbb{E}_\mu[c_{\mathrm{r}}(q^*)|t^i] \leq \mathbb{E}_\mu[c_{\mathrm{r}'}(q^*)|t^i], \quad \forall \mathrm{r}' \in R. \tag{4.12}$$

In a BWE, for each population type t^i, the *expected cost* as in (4.11) along any route taken by its travelers must be less than or equal to the expected cost of any of the other routes. Thus, any individual traveler in any population has no incentive to

deviate based on their belief when all travelers make route choices according to the equilibrium strategy.

The Objective BCG can be shown to be a convex, weighted potential game, where the equilibrium solution is found at the extremum of a weighted potential function [104]. This potential function can be easily solved analytically. In addition, due to the convexity of the weighted potential function, solutions can be found iteratively via gradient-descending including best-response dynamics methods such as fictitious play [105, 106].

4.4 Subjective Bayesian Congestion Game Formulation

In the Objective BCG, we adopted the common prior assumption, and as a result, the conditional type distribution of the other population's signal, $\Pr(t^{-i}|s)$, was common knowledge. As we have previously mentioned, this assumption is restrictive, since it assumes that populations correctly know the accuracy of the other populations information. This is unrealistic in many scenarios; for example, if we are trying to model a population of travelers who are entirely unaware of the existence of traveler information services, it does not make sense for them to know the accuracy of an information service they do not know exists. We contend that the common prior assumption that everyone knows the accuracy of everyone else's information is a very special case, and should only be used to model situations where there is plausible justification of such common knowledge. Instead, we believe that most real-world scenarios will incorporate some level of subjectivity, due to human limitations of knowledge, biases, and misconceptions.

We thus explore relaxing the common prior assumption by allowing populations to have inaccurate subjective beliefs of the opponent's information accuracy (also referred to as their *perceived accuracy* of their opponent's information). The Subjective BCG formulation is otherwise identical to the Objective BCG with one distinct dif-

ference: the conditional type distribution term is replaced with a subjective belief of the conditional type distribution, $\Pr^i(t^{-i}|s)$, for each player population.

4.4.1 Perceived Accuracy and Subjective Beliefs

To specifically explore the effect of subjective perceived accuracy, we consider a stylized case where population H receives perfectly accurate information about the state, and Population L receives no information, i.e.

$$\Pr(t^{\mathrm{H}} = s|s) = 1, \quad \Pr(t^{\mathrm{H}} \neq s|s) = 0, \quad \forall s \in S, \tag{4.13a}$$

$$\Pr(t^{\mathrm{L}} = s|s) = 0.5, \quad \Pr(t^{\mathrm{L}} \neq s|s) = 0.5, \quad \forall s \in S. \tag{4.13b}$$

We denote population L's perceived accuracy of t^{H} as η^{H}, and population H's subjective belief of the accuracy of t^{L} as η^{L}. For each population i, $\eta^i \in [0.5, 1]$. This parameter quantifies the belief of population $-i$ on how much information does population i has. If $\eta^i = 0.5$ (resp. $\eta^i = 1$), then population $-i$ believes that population i receives no information (resp. receives complete information) of the state.

Analogous to (4.13), we can write population i's perception on the other population's conditional information distribution, denoted $\Pr^i(t^{-i}|s)$, as follows:

$$\Pr^{\mathrm{L}}(t^{\mathrm{H}} = s|s) = \eta^{\mathrm{H}}, \quad \Pr^{\mathrm{L}}(t^{\mathrm{H}} \neq s|s) = 1 - \eta^{\mathrm{H}}, \quad \forall s \in S, \tag{4.14a}$$

$$\Pr^{\mathrm{H}}(t^{\mathrm{L}} = s|s) = \eta^{\mathrm{L}}, \quad \Pr^{\mathrm{H}}(t^{\mathrm{L}} \neq s|s) = 1 - \eta^{\mathrm{L}}, \quad \forall s \in S. \tag{4.14b}$$

Indeed, the case that each population correctly knows the opponent's information accuracy is a special case in our model, i.e. $\eta^{\mathrm{H}} = 1$ and $\eta^{\mathrm{L}} = 0.5$. We refer to this case as the *objective* belief case, as the belief of the conditional information distribution in (4.14) is identical to the true distribution in (4.4).

The subjective variant of the belief given in (4.5) with the perceived accuracies is

given as

$$\tilde{\mu}^i(s,t^{-i}|t^i) = \frac{\theta(s)\Pr(t^i|s)}{\Pr(t^i)} \cdot \Pr^i(t^{-i}|s), \quad \forall s \in s, \quad \forall t^i \in T^i, \quad \forall t^{-i} \in T^{-i}, \quad \forall i \in I. \tag{4.15}$$

From equations (4.13), (4.14) and (4.15), we calculate the expressions for $\tilde{\mu}^i(s,t^{-i}|t^i)$ for this stylized configuration of the game in Table 4.1.

	$s=\mathrm{a}$	$s=\mathrm{n}$
$t^{\mathrm{L}}=\mathrm{a}$	η^{L}	0
$t^{\mathrm{L}}=\mathrm{n}$	$1-\eta^{\mathrm{L}}$	0

(a) $\tilde{\mu}^{\mathrm{H}}(s,t^{\mathrm{L}}|t^{\mathrm{H}}=\mathrm{a})$

	$s=\mathrm{a}$	$s=\mathrm{n}$
$t^{\mathrm{L}}=\mathrm{a}$	0	$1-\eta^{\mathrm{L}}$
$t^{\mathrm{L}}=\mathrm{n}$	0	η^{L}

(b) $\tilde{\mu}^{\mathrm{H}}(s,t^{\mathrm{L}}|t^{\mathrm{H}}=\mathrm{n})$

	$s=\mathrm{a}$	$s=\mathrm{n}$
$t^{\mathrm{H}}=\mathrm{a}$	$p\eta^{\mathrm{H}}$	$(1-p)(1-\eta^{\mathrm{H}})$
$t^{\mathrm{H}}=\mathrm{n}$	$p(1-\eta^{\mathrm{H}})$	$(1-p)\eta^{\mathrm{H}}$

(c) $\tilde{\mu}^{\mathrm{L}}(s,t^{\mathrm{L}}|t^{\mathrm{H}}=\mathrm{a}) = \mu^{\mathrm{L}}(s,t^{\mathrm{L}}|t^{\mathrm{H}}=\mathrm{n})$

Table 4.1: Belief distributions of each population

For comparison, the beliefs of the objective case are also given for the stylized case. To distinguish the objective case from the general subjective belief structure, we denote $\bar{\mu}^i(s,t^{-i}|t^i)$ as the objective belief. By inserting $\eta^{\mathrm{H}}=1$ and $\eta^{\mathrm{L}}=0.5$ into the expressions in Table 4.1, $\bar{\mu}^i(s,t^{-i}|t^i)$ can be simplified as follows:

$$\bar{\mu}^{\mathrm{H}}(s,t^{\mathrm{L}}|t^{\mathrm{H}}) = \begin{cases} 0.5, & \text{if } s=t^{\mathrm{H}}, \\ 0, & \text{o.w.}, \end{cases} \quad \forall t^{\mathrm{L}} \in T^{\mathrm{L}}, \quad \forall t^{\mathrm{H}} \in T^{\mathrm{H}},$$

$$\bar{\mu}^{\mathrm{L}}(s,t^{\mathrm{H}}|t^{\mathrm{L}}) = \begin{cases} \theta(s), & \text{if } s=t^{\mathrm{H}}, \\ 0, & \text{o.w.}, \end{cases} \quad \forall t^{\mathrm{L}} \in T^{\mathrm{L}}, \quad \forall t^{\mathrm{H}} \in T^{\mathrm{H}}.$$

These expressions will be used to derive the equilibrium route choices when we solve the game in the following section.

Chapter 5

Equilibrium Analysis

This chapter presents the equilibrium route choices and costs for the Subjective BCG formulation discussed in the previous chapter. Recall that the Objective BCG is a special case of the Subjective BCG, and thus the solution to that game is contained within the analysis of the Subjective BCG.

We first present the route choices in Sec. 5.1, and provide analytic expressions for the Bayesian Wardrop Equilibrium route choices for each population and type. We characterize the equilibrium route choices with respect to the fraction of informed population, perceived accuracies, and probability of incident. The solutions are divided in three classes of solutions, based on the probability of incident and a "misperception factor" which characterizes the effect of subjectivity. The solution for one of the classes (dubbed "Class O") is identical to the solution to the Objective BCG. Within each class, there are a number of qualitatively distinct regimes depending on the fraction of informed population and the perceived accuracy.

In Sec. 5.2, we present an analysis of the costs for each equilibrium class and regime. We study how the costs for each population and all travelers on average are affected by each of the parameters to characterize the value of information. We demonstrate that there is a diminishing value of information to those receiving accurate information as more travelers gain access to information, eventually reaching zero. We also demonstrate that the perception of information accuracy has a significant effect on equilibrium costs, and under certain conditions of incident probability

and fraction of travelers with information, it is beneficial for those with access to accurate information to pretend that the information they have is less accurate than it actually is; yet, for other configurations, it is beneficial to be honest about the accuracy of their information.

Finally, in Sec. 5.3, we show that the Subjective BCG can be solved in general as a finite set of variational inequalities. This provides a solution method for the Subjective BCG, which, unlike the Objective BCG, does not admit a convex, weighted potential function. The variational inequality optimization problem is nonlinear and nonconvex, but the constraints are quadratic, and we were able to solve the game with different information structures using a Sequential Quadratic Programming nonlinear solver.

5.1 Equilibrium Route Choices

In this section, we analyze how the qualitative properties of equilibrium change with the probability of incident p, fraction of informed population λ^{H}, and the perceived accuracies η^{H}, η^{L}.

If in equilibrium, all travelers take the same route, then changing the population fraction λ^{H} does not change the equilibrium outcome. Therefore, to avoid triviality, we will assume that the total demand is sufficiently large that the cost of placing all of the demand on the first route in the normal state is costlier than the free-flow cost of the second route, i.e.:

$$\mathrm{D} > \frac{\beta_2 - \beta_1}{\alpha_1^{\mathrm{n}}}. \tag{5.1}$$

Hence, there is no equilibrium where all the demand is routed on the same route.

We first present some preliminary results that are useful for our subsequent analysis in Sec. 5.1.1, and then provide complete equilibrium characterization in Sec. 5.1.2.

5.1.1 Preliminary results

We first show that the type t^{L} does not impact population L's strategy. As a consequence, the perceived accuracy of population L's information has no impact on the equilibrium.

Lemma 1. *The equilibrium strategy of population* L *is independent of the type, i.e.* $q^{L*}(\mathrm{a}) = q^{L*}(\mathrm{n})$. *Additionally, the equilibrium strategy profile* $\left(q^{H*}, q^{L*}\right)$ *does not change with* η^{L}.

The intuition of this result is that since the beliefs $\mu^{\mathrm{L}}(s, t^{\mathrm{H}}|t^{\mathrm{L}})$ are identical for both $t^{\mathrm{L}} \in \{\mathrm{a}, \mathrm{n}\}$ (Table 4.1), the expected cost of each route is independent of the type. Therefore, the equilibrium strategy of population L as defined in Definition 1 is also independent of the type. Consequently, population H's perception on population L's accuracy does not change the equilibrium strategies. We can thus drop η^{L} from our further analysis. Additionally, we simplify our notation by defining $q^{\mathrm{Ha}*} \triangleq q^{H*}(\mathrm{a})$, $q^{\mathrm{Hn}*} = q^{H*}(\mathrm{n})$, and $q^{L*} \triangleq q^{L*}(\mathrm{a}) = q^{L*}(\mathrm{n})$. We denote the aggregate flow on route r as $l_r^{\mathrm{a}*}$ (resp. $l_{\mathrm{r}}^{\mathrm{n}*}$) when the state is a (resp. n).

Each population, given the information it receives, can either assign all its demand on r_1 (i.e. $q_1^i(t^i) = \lambda^i \mathrm{D}$), splits on both routes (i.e. $q_1^i(t^i) \in (0, \lambda^i \mathrm{D})$), or on r_2 (i.e. $q_1^i(t^i) = 0$). Thus, there are in total $3 \times 3 \times 3 = 27$ possible cases for the equilibrium strategy profile $q^* = \left(q^{\mathrm{Ha}*}, q^{\mathrm{Hn}*}, q^{L*}\right)$.

The following proposition provides three properties that equilibrium strategies must satisfy. These properties help us rule out some cases that cannot be equilibrium.

Proposition 1. *Any equilibrium strategy profile must satisfy the following properties:*

(i) *If* $q_1^{\mathrm{Hn}*} = \lambda^{\mathrm{H}} \mathrm{D}$, *then* $q_1^{L*} < (1 - \lambda^{\mathrm{H}})\mathrm{D}$. *If* $q_1^{\mathrm{Ha}*} = 0$, *then* $q_1^{L*} > 0$.

(ii) $q_1^{\mathrm{Ha}*} \leq q_1^{\mathrm{Hn}*}$.

(iii) $q_1^{\mathrm{Ha}*} < \lambda^{\mathrm{H}} \mathrm{D}$, *and* $q_1^{\mathrm{Hn}*} > 0$.

Among all 27 cases, only 8 cases satisfy properties (i) – (iii), see Table 5.1. We now provide interpretation and intuition for each property:

(i) This property excludes the case that all travelers exclusively takes one route in equilibrium as the demand satisfies (5.1).

(ii) This property ensures that the amount of travelers in population H who choose r_1 in state a is no higher than that in state n. Assume for the sake of contradiction that more travelers in population H take r_1 in state a than that in state n. Then the cost of r_1 in state a must be strictly higher than that in state n. This cannot be an equilibrium since travelers will deviate to take r_2 in state a.

(iii) This property excludes the case that population H takes r_1 exclusively in state a, and r_2 in state n. The intuition is that if Ha takes r_1 exclusively, then by comparing the expected costs of both routes based on the beliefs of Hn and L, we can show that both Hn and L will take r_1 exclusively. This is a contradiction since when D satisfies (5.1), travelers will not only take one routes in equilibrium. Analogously, if Hn takes r_2 exclusively, then we can show that all travelers must take r_2 exclusively, which cannot be in equilibrium.

Ha \ Hn	r_1	split	r_2
r_1	(iii)	(ii)	(ii)
split	(i)	✓	(ii)
r_2	(i)	✓	(iii)

(a) L takes r_1

Ha \ Hn	r_1	split	r_2
r_1	(iii)	(ii)	(ii)
split	✓	✓	(ii)
r_2	✓	✓	(iii)

(b) L splits

Ha \ Hn	r_1	split	r_2
r_1	(iii)	(ii)	(ii)
split	✓	✓	(ii)
r_2	(i)	(i)	(iii)

(c) L takes r_2

Table 5.1: Equilibrium Strategy profiles in Bayesian routing game. Possible cases are marked with a check. Non-equilibrium cases are marked with the violated property.

5.1.2 Equilibrium Characterization

Before presenting the equilibrium characterization, we first introduce the following key quantities that are used throughout our analysis:

$$K = \alpha_2 \mathrm{D} + \beta_2 - \beta_1, \quad A = \alpha_1^{\mathrm{a}} + \alpha_2, \quad N = \alpha_1^{\mathrm{n}} + \alpha_2, \quad \widehat{p} = \frac{N}{A+N}, \quad \rho = pA - (1-p)N. \tag{5.2}$$

We now interpret these quantities: The value of K is the cost difference between r_2 and r_1 when all the travelers take r_2. Since $\beta_2 > \beta_1$, $K > 0$. If all travelers are informed (i.e. $\lambda^H = 1$) and the state is a, then the costs of two routes must be identical in equilibrium, i.e. $c_1^a(l_1^{a*}) = c_2(D - l_1^{a*})$. The equilibrium flow on r_1 is $\frac{K}{A}$. Analogously, in state n, the equilibrium route flow on r_1 is $\frac{K}{N}$. Therefore, we can view A (resp. N) as the network congestion level when the state is a (resp. n). In complete information game, the equilibrium route flow on r_1 increases in the cost difference between r_2 and r_1 (i.e. K), and decreases in the network congestion level. Since the congestion level in state a is higher than that in state n (i.e. $A > N$), the demand on route r_1 in state a is less than that in state n.

Additionally, pA (resp. $(1-p)N$) is the network congestion level A (resp. N) weighted by the probability of the state a (resp. n). The value ρ is the difference of the weighted network congestion level between state a and n, and the threshold probability $\widehat{p}$ is the probability of state a such that the weighted network congestion level is the same in two states. We can check that $\widehat{p} < 0.5$, and $\widehat{p}$ decreases as the difference between the slopes of r_1 in the two states (i.e. $\alpha_1^a - \alpha_1^n$) increases. The value ρ is positive, zero, or negative, depending on whether p is higher, equal or lower than the threshold $\widehat{p}$:

$$\rho \triangleq pA - N(1-p) \begin{cases} > 0 & \text{if } p > \widehat{p}, \\ = 0 & \text{if } p = \widehat{p}, \\ < 0 & \text{if } p < \widehat{p}. \end{cases} \tag{5.3}$$

We say $(1-\eta^H)$ is the perception bias, and $\left(1-\eta^H\right)\rho$ is the misperception factor with the scaler ρ. The misperception factor is 0 if either $\eta^H = 1$ (i.e. objective belief case) or $\rho = 0$ (i.e. the scaler is 0 when $p = \widehat{p}$). From (5.3), we know that the misperception factor is positive if $p > \widehat{p}$, and negative if $p < \widehat{p}$.

We are now in the position to introduce the equilibrium regimes. Each regime corresponds to a range of λ^H such that the qualitative properties of equilibrium (i.e., the routes taken by travelers in each population) do not change in the interior of each

regime.

It turns out that for a given λ^{H}, the qualitative properties of the equilibrium strategies can be different depending on whether the misperception factor $\rho(1-\eta^{\mathrm{H}})$ is positive, zero or negative. Therefore, we define three classes of regimes: Class I regimes ($\rho(1-\eta^{\mathrm{H}})>0$), Class II regimes ($\rho(1-\eta^{\mathrm{H}})<0$), and Class O ("Effectively objective class" $\rho(1-\eta^{\mathrm{H}})=0$):

- Class I regimes ($\rho(1-\eta^{\mathrm{H}})>0$): Four equilibrium regimes are ϕ^1: $\lambda^{\mathrm{H}}\in[0,\Lambda^1)$, ϕ^2: $\lambda^{\mathrm{H}}\in(\Lambda^1,\Lambda^2)$, ϕ^3: $\lambda^{\mathrm{H}}\in[\Lambda^2,\Lambda^3)$, and ϕ^4: $\lambda^{\mathrm{H}}\in(\Lambda^3,1)$. The thresholds $\Lambda^1-\Lambda^3$ are in (5.4a).

- Class II regimes ($\rho(1-\eta^{\mathrm{H}})<0$): Four equilibrium regimes are ϕ_1: $\lambda^{\mathrm{H}}\in(0,\Lambda_1)$, ϕ_2: $\lambda^{\mathrm{H}}\in[\Lambda_1,\Lambda_2)$, ϕ_3: $\lambda^{\mathrm{H}}\in[\Lambda_2,\Lambda_3)$, and ϕ_4: $\lambda^{\mathrm{H}}\in(\Lambda_3,1)$. The thresholds $\Lambda_1-\Lambda_3$ are in (5.4b).

- Class O regimes ($\rho(1-\eta^{\mathrm{H}})=0$): Two equilibrium regimes are $\bar{\phi}_1$: $\lambda^{\mathrm{H}}\in(0,\bar{\Lambda})$, $\bar{\phi}_2$: $\lambda^{\mathrm{H}}\in[\bar{\Lambda},1)$. The threshold $\bar{\Lambda}$ is in (5.4c).

$$\Lambda^1=\frac{K}{A\mathrm{D}}\frac{(1-p)(A-N)}{(1-p)N+\rho(1-\eta^{\mathrm{H}})},\ \Lambda^2=\frac{K}{A\mathrm{D}}\left(1+\frac{(1-p)(A-N)}{(1-p)N+\rho(1-\eta^{\mathrm{H}})}\right),\ \Lambda^3=\frac{K}{N\mathrm{D}}, \tag{5.4a}$$

$$\Lambda_1=\frac{K}{N\mathrm{D}}\frac{p(A-N)}{pA-\rho(1-\eta^{\mathrm{H}})},\qquad \Lambda_2=1-\frac{K}{N\mathrm{D}}+\frac{K}{N\mathrm{D}}\frac{p(A-N)}{pA-\rho(1-\eta^{\mathrm{H}})},\qquad \Lambda_3=1-\frac{K}{A\mathrm{D}}. \tag{5.4b}$$

$$\bar{\Lambda}=\frac{K}{\mathrm{D}}\left(\frac{1}{N}-\frac{1}{A}\right). \tag{5.4c}$$

Lemma 2 presents the properties of the regime thresholds.

Lemma 2. *If* $\rho(1-\eta^{\mathrm{H}})>0$, *then the thresholds of class I regimes satisfy* $0<\Lambda^1<\Lambda^2<\Lambda^3<1$. *If* $\rho(1-\eta^{\mathrm{H}})<0$, *then the thresholds of class II regimes satisfy* $0<\Lambda_1<\Lambda_2<\Lambda_3<1$. *In class O regimes, the threshold satisfies* $0<\bar{\Lambda}<1$. *Additionally,*

as $\rho(1-\eta^{\mathrm{H}})$ *goes to 0, the regime thresholds in* (5.4a) – (5.4b) *can be simplified as follows:*

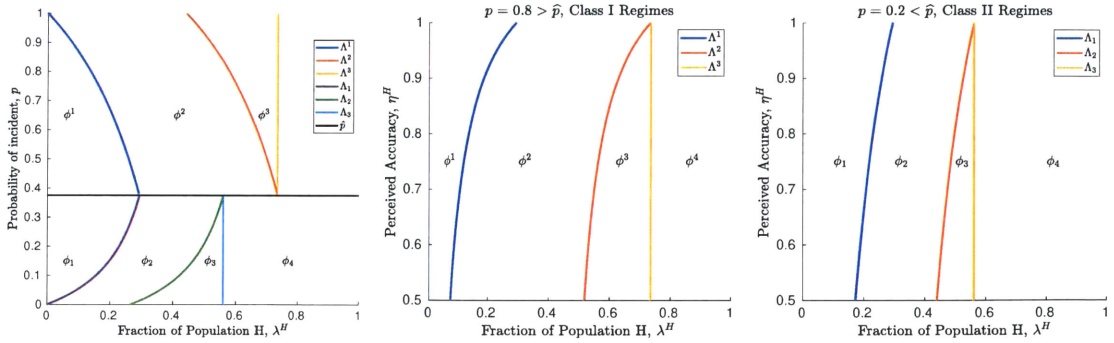
$$\lim_{\rho(1-\eta^{\mathrm{H}})\to 0} \Lambda^1 = \lim_{\rho(1-\eta^{\mathrm{H}})\to 0} \Lambda_1 = \bar{\Lambda}, \qquad \lim_{\rho(1-\eta^{\mathrm{H}})\to 0} \Lambda^2 = \Lambda^3, \qquad \lim_{\rho(1-\eta^{\mathrm{H}})\to 0} \Lambda_2 = \Lambda_3.$$

That is, as the misperception factor $\rho(1-\eta^{\mathrm{H}})$ goes to zero, regime ϕ^1 in class I (resp. ϕ_1 in class II) is equivalent to the regime $\bar{\phi}_1$ in the equivalently objective class, and regime ϕ^3 in class I (resp. ϕ_3 in class II) becomes empty. Regimes ϕ^2 and ϕ^4 in class I (resp. ϕ_2 and ϕ_4 in class II) are merged to be regime $\bar{\phi}_2$ in class O.

We illustrate how regime boundaries change with η^{H} and λ^{H} in Figure 5-1. Unless otherwise specified, the parameters for the cost functions of the two routes in our numerical examples take the values given in Table 5.2.

Quality	Symbol	Value
r_1 slope, state n	α_1^{n}	1
r_1 slope, state a	α_1^{a}	3
r_2 slope	α_2	2
r_1 intercept	β_1	1
r_2 intercept	β_2	1.2
Total demand	D	1

Table 5.2: Parameter values for two route network example.



(a) (b) (c)

Figure 5-1: (a) Regimes and boundaries with $\eta^{\mathrm{H}} = 0.8$. The horizontal line at $\hat{p}$ divides Class I regimes (top) from Class II regimes (bottom); (b) Class I regimes with $p = 0.8$; (c) Class II regimes with $p = 0.2$.

The next theorem presents the complete equilibrium characterization in each

regime.

Theorem 1. *Class I regimes ($\rho(1-\eta^{\mathrm{H}})>0$): equilibrium strategy profile is unique in each regime.*

- *Regime ϕ^1:* $\lambda^{\mathrm{H}} \in (0, \Lambda^1)$.

$$q_1^{\mathrm{H}*}(\mathrm{a}) = 0, \quad q_1^{\mathrm{H}*}(\mathrm{n}) = \lambda^{\mathrm{H}}\mathrm{D}, \quad q_1^{\mathrm{L}*} = \frac{K - \left(N(1-p) + \rho(1-\eta^{\mathrm{H}})\right)\lambda^{\mathrm{H}}\mathrm{D}}{pA + (1-p)N}. \tag{5.5}$$

- *Regime ϕ^2:* $\lambda^{\mathrm{H}} \in [\Lambda^1, \Lambda^2)$.

$$q_1^{\mathrm{H}*}(\mathrm{a}) = \frac{K}{A} - q_1^{\mathrm{L}*}, \quad q_1^{\mathrm{H}*}(\mathrm{n}) = \lambda^{\mathrm{H}}\mathrm{D}, \quad q_1^{\mathrm{L}*} = \Lambda^2\mathrm{D} - \lambda^{\mathrm{H}}\mathrm{D}.$$

- *Regime ϕ^3:* $\lambda^{\mathrm{H}} \in [\Lambda^2, \Lambda^3)$.

$$q_1^{\mathrm{H}*}(\mathrm{a}) = \frac{K}{A}, \quad q_1^{\mathrm{H}*}(\mathrm{n}) = \lambda^{\mathrm{H}}\mathrm{D}, \quad q_1^{\mathrm{L}*} = 0.$$

- *Regime ϕ^4:* $\lambda^{\mathrm{H}} \in [\Lambda^3, 1)$.

$$q_1^{\mathrm{H}*}(\mathrm{a}) = \frac{K}{A}, \quad q_1^{\mathrm{H}*}(\mathrm{n}) = \frac{K}{N}, \quad q_1^{\mathrm{L}*} = 0.$$

Class II regimes ($\rho(1-\eta^{\mathrm{H}})<0$): equilibrium strategy profile is unique in each regime.

- *Regime ϕ_1:* $\lambda^{\mathrm{H}} \in (0, \Lambda_1)$. *The equilibrium strategy profile* $\left(q_1^{\mathrm{H}*}(\mathrm{a}), q_1^{\mathrm{H}*}(\mathrm{n}), q_1^{\mathrm{L}*}\right)$ *is identical to that in regime ϕ^1.*

- *Regime ϕ_2:* $\lambda^{\mathrm{H}} \in [\Lambda_1, \Lambda_2)$.

$$q_1^{\mathrm{H}*}(\mathrm{a}) = 0, \quad q_1^{\mathrm{H}*}(\mathrm{n}) = \frac{K}{N} - q_1^{\mathrm{L}*}, \quad q_1^{\mathrm{L}*} = (1-\Lambda_2)\mathrm{D}.$$

- *Regime ϕ_3:* $\lambda^{\mathrm{H}} \in [\Lambda_2, \Lambda_3)$.

$$q_1^{\mathrm{H}*}(\mathrm{a}) = 0, \quad q_1^{\mathrm{H}*}(\mathrm{n}) = \frac{K}{N} - (1-\lambda^{\mathrm{H}})\mathrm{D}, \quad q_1^{\mathrm{L}*} = (1-\lambda^{\mathrm{H}})\mathrm{D}.$$

- *Regime ϕ_4:* $\lambda^{\mathrm{H}} \in [\Lambda_3, 1)$.

$$q_1^{\mathrm{H}*}(\mathrm{a}) = \frac{K}{A} - (1-\lambda^{\mathrm{H}})\mathrm{D}, \quad q_1^{\mathrm{H}*}(\mathrm{n}) = \frac{K}{N} - (1-\lambda^{\mathrm{H}})\mathrm{D}, \quad q_1^{\mathrm{L}*} = (1-\lambda^{\mathrm{H}})\mathrm{D}.$$

Class O regimes ($\rho(1-\eta^{\mathrm{H}}) = 0$): equilibrium strategy profile is unique in regime $\bar{\phi}_1$, but not in regime $\bar{\phi}_2$.

- *Regime $\bar{\phi}_1$:* $\lambda^{\mathrm{H}} \in (0, \bar{\Lambda})$. *The unique equilibrium strategy profile is identical to that in regime ϕ^1, which can be simplified as follows:*

$$q_1^{H*}(\mathrm{a}) = 0, \quad q_1^{H*}(\mathrm{n}) = \lambda^{\mathrm{H}}\mathrm{D}, \quad q_1^{L*} = \frac{K - \lambda^{\mathrm{H}}\mathrm{D}(1-p)N}{pA + (1-p)N}.$$

- *Regime $\bar{\phi}_2$:* $\lambda^{\mathrm{H}} \in [\bar{\Lambda}, 1)$. *The set of equilibrium strategy profiles can be written as:*

$$q_1^{H*}(\mathrm{a}) = \gamma, \quad q_1^{H*}(\mathrm{n}) = \bar{\Lambda}\mathrm{D} + \gamma, \quad q_1^{L*} = \frac{K}{A} - \gamma, \tag{5.6}$$

where $\max\{0, \frac{K}{A} - (1-\lambda^{\mathrm{H}})\mathrm{D}\} \leq \gamma \leq \min\{\frac{K}{A}, \lambda^{\mathrm{H}}\mathrm{D} - \bar{\Lambda}\mathrm{D}\}$.

We now discuss the qualitative properties of equilibrium strategy profile in each regime.

First, we emphasize that although the equilibrium strategy profile is not unique in regime $\bar{\phi}_2$ of Class O, it is *essentially unique* in that the aggregate equilibrium flow on r_1 is $\frac{K}{A}$ in state a and $\frac{K}{N}$ in state n. The equilibrium flow is identical to that in regimes ϕ^4 of class I and ϕ_4 of class II. In this regime, the expected costs of two routes are identical for both population H and L. We can check that for any $\max\{0, \frac{K}{A} - (1-\lambda^{\mathrm{H}})\mathrm{D}\} < \gamma < \min\{\frac{K}{A}, \lambda^{\mathrm{H}}\mathrm{D} - \bar{\Lambda}\mathrm{D}\}$, the corresponding equilibrium strategy in (5.6) is such that Ha, Hn and L all split on both routes.

Next, recall from Table 5.1 that there are only 8 cases that are possible in equilibrium. Theorem 1 shows that in each regime, the equilibrium strategy profile falls into one of the 8 cases, which is summarized in Table 5.3.

	ϕ^1	ϕ^2	ϕ^3	ϕ^4
Ha	r_2	split	split	split
Hn	r_1	r_1	r_1	split
L	split	split	r_2	r_2

(a) Class I regimes

	ϕ^1	ϕ^2	ϕ^3	ϕ^4
Ha	r_2	r_2	r_2	split
Hn	r_1	split	split	split
L	split	split	r_1	r_1

(b) Class II regimes

	$\bar{\phi}_1$	$\bar{\phi}_2$
Ha	r_2	split
Hn	r_1	split
L	split	split

(c) Class O regimes

Table 5.3: Equilibrium strategy profiles in Bayesian routing game.

In all three classes of regimes, as λ^H increases, the change of population H's strategy is similar. When λ^H is small (i.e. in regimes ϕ^1, ϕ_1 and $\bar{\phi}_1$), population H exclusively takes r_2 in state a, and r_1 in state n. As λ^H increases, population H can no longer take one route exclusively in both states due to the increasing congestion externalities caused by travelers in the same population. Therefore, in each state, population H starts to split on both routes when the population size λ^H is above a certain threshold. Eventually, when the majority of travelers are informed (regimes ϕ^4, ϕ_4, and $\bar{\phi}_2$), travelers in H split on both routes in both states.

The equilibrium strategies in the three classes of regimes are different from two aspects. First, the thresholds of λ^H beyond which population H starts to split on both routes are different: In class I regimes, the threshold that Ha starts to splits on both routes (Λ^1) is smaller than the threshold for Hn (Λ^3). Therefore, when λ^H is relatively medium ($\lambda^H \in [\Lambda^1, \Lambda^3]$), H splits in state a, but takes r_1 exclusively in state n; In class II regimes, the threshold for Ha (Λ_3) is higher than the threshold for Hn (Λ_1). Therefore, when λ^H is relatively medium ($\lambda^H \in [\Lambda_1, \Lambda_3]$), H splits in state n, but takes r_2 exclusively in state a; In class O regimes, the thresholds for Ha and Hn are the same ($\bar{\Lambda}$). In both states, population H either takes one route exclusively or splits on both routes.

Second, the perception bias $(1-\eta^H)$ impacts population L's strategy differently in the three classes of regimes. For any $\eta^H < 1$, if the probability of state a is higher than $\widehat{p}$, i.e. class I (resp. lower than $\widehat{p}$, i.e. class II), then the demand of travelers in population L who take r_1 is lower (resp. higher) than that in the case of objective belief ($\eta^H = 1$) for any given λ^H, and the difference is scaled by the absolute scaler $|\rho|$. If $p = \widehat{p}$ (i.e. class O), then the equilibrium strategy is identical to the case with

objective belief. Specifically, when λ^{H} is high, population L exclusively takes r_2 for $\lambda^{\mathrm{H}} > \Lambda^2$ in class I, takes r_1 for $\lambda^{\mathrm{H}} > \Lambda_2$ in class II, and splits on both routes for $\lambda^{\mathrm{H}} > \bar{\Lambda}$ in class O.

The reason for these differences is that the misperception factor impacts population L's expectation on the costs of two routes differently in the three classes. We calculate the cost difference between the two routes in equilibrium for any state s and any type of population H:

$$\begin{aligned}\Delta^s(l^{t^{\mathrm{H}}*}) &\triangleq c_1^s\left(l_1^{t^{\mathrm{H}}*}\right) - c_2\left(\mathrm{D} - l_1^{t^{\mathrm{H}}*}\right) \\ &= \begin{cases} c_1^{\mathrm{a}}(l_1^{\mathrm{a}*}) - c_2(\mathrm{D} - l_1^{\mathrm{a}*}) = Al_1^{\mathrm{a}*} - K, & \text{if } s = \mathrm{a} \text{ and } t^{\mathrm{H}} = \mathrm{a}, \\ c_1^{\mathrm{a}}(l_1^{\mathrm{n}*}) - c_2(\mathrm{D} - l_1^{\mathrm{n}*}) = Al_1^{\mathrm{n}*} - K, & \text{if } s = \mathrm{a} \text{ and } t^{\mathrm{H}} = \mathrm{n}, \\ c_1^{\mathrm{n}}(l_1^{\mathrm{a}*}) - c_2(\mathrm{D} - l_1^{\mathrm{a}*}) = Nl_1^{\mathrm{a}*} - K, & \text{if } s = \mathrm{n} \text{ and } t^{\mathrm{H}} = \mathrm{a}, \\ c_1^{\mathrm{n}}(l_1^{\mathrm{n}*}) - c_2(\mathrm{D} - l_1^{\mathrm{n}*}) = Nl_1^{\mathrm{n}*} - K, & \text{if } s = \mathrm{n} \text{ and } t^{\mathrm{H}} = \mathrm{n}, \end{cases} \end{aligned} \tag{5.7}$$

Recall from Proposition 1 $q_1^{\mathrm{Ha}*} \leq q_1^{\mathrm{Hn}*}$, hence $l_1^{\mathrm{a}*} \leq l_1^{\mathrm{n}*}$. If $\eta^{\mathrm{H}} = 1$ (the case of objective beliefs), then the expected cost difference between the two routes is $\mathbb{E}_{\bar{\mu}}[\Delta^s(l^*)] = p\Delta^{\mathrm{a}}(l_1^{\mathrm{a}*}) + (1-p)\Delta^{\mathrm{n}}(l_1^{\mathrm{n}*})$. With perceived accuracy $\eta^{\mathrm{H}} < 1$, if population L believes that population H receive the correct information of the state, then the expected cost difference is identical to the case with objective belief. However, if the state is a but population L believes that population H receives the the information n ($s = \mathrm{a}$, $t^{\mathrm{L}} = \mathrm{n}$), then population L *over-estimates* the cost difference by $A(l_1^{\mathrm{n}*} - l_1^{\mathrm{a}*})$ compared with the case with objective belief ($s = t^{\mathrm{L}} = \mathrm{a}$). If the state is n but population L believes that population H receives the the information a ($s = \mathrm{n}$, $t^{\mathrm{L}} = \mathrm{a}$), then population L *under-estimates* the cost of r_1 by $N(l_1^{\mathrm{n}*} - l_1^{\mathrm{a}*})$. Recall population L's belief in Table 4.1c, we obtain:

$$\begin{aligned}&\mathbb{E}_{\mu}[\Delta^s(l^*)|\mathrm{L}] - \mathbb{E}_{\bar{\mu}}[\Delta^s(l^*)|\mathrm{L}] = p(1-\eta^{\mathrm{H}})A(l_1^{\mathrm{n}*} - l_1^{\mathrm{a}*}) - (1-p)(1-\eta^{\mathrm{H}})N(l_1^{\mathrm{n}*} - l_1^{\mathrm{a}*}) \\ =&\rho(1-\eta^{\mathrm{H}})(l_1^{\mathrm{n}*} - l_1^{\mathrm{a}*}) \begin{cases} > 0, & \text{in Class I}, \\ = 0, & \text{in Class O}, \\ < 0, & \text{in Class II}. \end{cases} \end{aligned} \tag{5.8}$$

If the misperception factor $\rho(1-\eta^{\mathrm{H}})$ is zero, then the expected costs difference between the two routes are identical to that in objective belief case because either population L has objective belief $(1-\eta^{\mathrm{H}}=0)$, or the over-estimation of cost difference in state a and the under-estimation in state n cancel out $(\rho=0)$. Therefore, we say class O is effectively objective.

In class O, if $\lambda^{\mathrm{H}}=0$, then the expected network congestion level for population L is $pA+(1-p)N$. Therefore, the demand in population L on r_1 is $\frac{K}{pA+(1-p)N}$. As λ^{H} increases in regime $\bar{\phi}_1$, H takes r_2 in state a since $c_1^{\mathrm{a}}(l_1^{\mathrm{a}*})-c_2(\mathrm{D}-l_1^{\mathrm{a}*})=Al_1^{\mathrm{a}*}-K>0$ (i.e. $l_1^{\mathrm{a}*}>\frac{K}{A}$), and population H takes r_1 in state n since $c_1^{\mathrm{n}}(l_1^{\mathrm{n}*})-c_2(\mathrm{D}-l_1^{\mathrm{n}*})=Nl_1^{\mathrm{n}*}-K<0$ (i.e. $l_1^{\mathrm{n}*}<\frac{K}{A}$). As λ^{H} increases, the demand in population L on r_1 decreases with the rate $\frac{(1-p)N\mathrm{D}}{pA+(1-p)N}$. Therefore, equilibrium flow $l_1^{\mathrm{a}*}$ also decreases with the same rate, and $l_1^{\mathrm{n}*}$ increases in λ^{H} with the rate $\frac{pA\mathrm{D}}{pA+(1-p)N}$. We can check that when λ^{H} reaches $\bar{\Lambda}$, $l_1^{\mathrm{a}*}$ decreases to $\frac{K}{A}$, and $l_1^{\mathrm{n}*}$ increases to $\frac{K}{N}$. Consequently, the costs of the two routes are identical for population H in both states. Hence, Ha and Hn both start to split at the threshold $\bar{\Lambda}$. From (5.7), the expected cost difference for population L becomes:

$$\mathbb{E}_{\mu}[\Delta^s(l^*)|\mathrm{L}]=\frac{K(A-N)}{AN}(1-\eta^{\mathrm{H}})\rho, \tag{5.9}$$

which is zero in $\bar{\phi}_2$. Therefore, population L also splits on both routes.

On the other hand, since the misperception factor $\rho(1-\eta^{\mathrm{H}})>0$ in class I, the over-estimation of the cost difference in state a is higher than the under-estimate of that in state n, i.e. $\mathbb{E}_{\mu}[\Delta^s(l^*)|\mathrm{L}]>\mathbb{E}_{\bar{\mu}}[\Delta^s(l^*)|\mathrm{L}]$ as in (5.8). Hence, q_1^{L*} decreases with a faster rate with λ^{H} in regime ϕ^1 than the rate in regime $\bar{\phi}_1$. As a consequence, $l_1^{\mathrm{a}*}$ decrease to $\frac{K}{A}$ when $\lambda^{\mathrm{H}}=\Lambda^1$, but $l_1^{\mathrm{n}*}$ is still smaller than $\frac{K}{N}$ until λ^{H} reaches a higher threshold Λ^3. For $\lambda^{\mathrm{H}}\in[\Lambda^1,\Lambda^3]$, Ha starts to split, but Hn takes r_1 exclusively. The demand of travelers in population L on route r_1 continues to decrease in λ^{H} (regime ϕ^2) until it reaches zero at Λ^2 (regime ϕ^3). In regime ϕ^4, population H splits in both states. The expected cost difference based on population L's belief is as in (5.9), which is positive. Hence, population L exclusively takes r_2.

Analogously, since the misperception factor $\rho(1-\eta^{\mathrm{H}}) < 0$ in class II, the over-estimation of the cost difference in state a is lower than the under-estimate of that in state n, i.e. $\mathbb{E}_{\mu}[\Delta^s(l^*)|\mathrm{L}] < \mathbb{E}_{\bar{\mu}}[\Delta^s(l^*)|\mathrm{L}]$. Hence, q_1^{L*} decreases with a slower rate with λ^{H} in regime ϕ^1. Consequently, $l_1^{\mathrm{n}*}$ increases to $\frac{K}{N}$ when $\lambda^{\mathrm{H}} = \Lambda_1$, but $l_1^{\mathrm{a}*}$ is still higher than $\frac{K}{A}$ until λ^{H} reaches a higher threshold Λ_3. For $\lambda^{\mathrm{H}} \in [\Lambda_1, \Lambda_3]$, Hn starts to split, but Ha takes r_2 exclusively. The demand of travelers in population L on route r_2 continues to decrease in λ^{H} (regime ϕ_2) until it reaches zero at Λ_2 (regime ϕ_3). In regime ϕ_4, population H splits in both states. The expected cost difference based on population L's belief in (5.9) is negative. Hence, population L exclusively takes r_1

5.2 Equilibrium Costs

In this section, we first study the cost difference between the two populations in each regime. Next, we analyze how the perceived accuracy of population H's information (η^{H}) impacts the equilibrium cost of travelers in each population, and the average cost of all travelers.

The equilibrium population cost C^{i*} is defined as the average cost experienced by a traveler of population i in equilibrium:

$$C^{\mathrm{H}*} \triangleq \frac{1}{\lambda^{\mathrm{H}}\mathrm{D}} \sum_{s\in S} \theta(s) \left(\sum_{\mathrm{r}\in R} c_r^s(q_r^{H*}(s) + q_r^{L*}) \cdot q_r^{H*}(s) \right), \tag{5.10a}$$

$$C^{\mathrm{L}*} \triangleq \frac{1}{(1-\lambda^{\mathrm{H}})\mathrm{D}} \sum_{s\in S} \theta(s) \left(\sum_{\mathrm{r}\in R} c_r^s(q_r^{H*}(s) + q_r^{L*}) \cdot q_r^{L*} \right). \tag{5.10b}$$

The average cost of all travelers in equilibrium, denoted C^* is defined as follows:

$$C^* \triangleq \lambda^{\mathrm{H}} C^{\mathrm{H}*} + (1-\lambda^{\mathrm{H}}) C^{\mathrm{L}*} = \frac{1}{\mathrm{D}} \sum_{s\in S} \theta(s) \sum_{\mathrm{r}\in R} c_r^s(l_r^{s*}) l_r^{s*}$$

5.2.1 Relative Value of Information

We define the relative value of information as the average travel cost saving that an informed traveler (population H) has compared with the uninformed traveler (popu-

lation L).

$$V^* \triangleq C^{\mathrm{L}*} - C^{\mathrm{H}*}.$$

The following proposition shows in each class, there exists a threshold of λ^{H} such that that the relative value of information is strictly positive if the fraction of the informed population is lower than the threshold, but zero if beyond.

Proposition 2. *In class I regimes,* $V^* > 0$ *if* $\lambda^{\mathrm{H}} < \Lambda^3$, *and* $V^* = 0$ *otherwise.*

In class II regimes, $V^* > 0$ *if* $\lambda^{\mathrm{H}} < \Lambda_3$, *and* $V^* = 0$ *otherwise.*

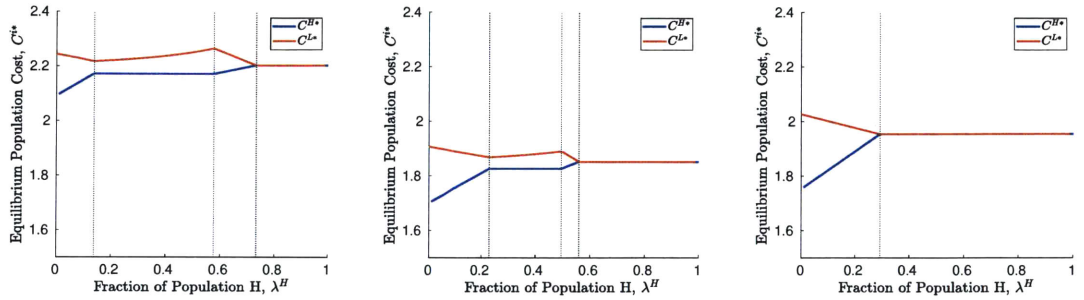
In class O regimes, $V^* > 0$ *if* $\lambda^{\mathrm{H}} < \bar{\Lambda}$, *and* $V^* = 0$ *otherwise.*

Proposition 2 shows that in all three classes of regimes, if the fraction of the informed travelers is below a certain threshold, then the informed travelers are better off than the uninformed on average. However, if most of the travelers are informed, then the two populations face identical average cost in equilibrium. From Theorem 1, when λ^{H} is relatively small, in one state or both states, population H takes the route with lower cost exclusively. However, a non-zero fraction of travelers in population L take the other route, which has a higher cost. Therefore, travelers in L face higher cost than travelers in H on average. However, when most of the travelers are informed, population H splits on both routes in both states due to the relatively large size. In this case, the costs of both routes are the same in both states. Hence, regardless of which route population L takes, travelers in L face identical cost with travelers in H.

The thresholds in the three classes are different. Specifically, both Λ^3 and Λ_3 are larger than $\bar{\Lambda}$. That is, if λ^{H} is relatively medium, i.e. $\lambda^{\mathrm{H}} \in [\bar{\Lambda}, \Lambda^3)$ in class I or $\lambda^{\mathrm{H}} \in [\bar{\Lambda}, \Lambda_3)$ in class II, then the relative value of information is positive when the misperception factor $\rho(1-\eta^{\mathrm{H}}) \neq 0$, but zero otherwise. Recall from (5.8), travelers in population L either over-estimates or under-estimates the expected cost difference between the two routes when the misperception factor $\rho(1-\eta^{\mathrm{H}})$ is non-zero. Population H can then leverage this biased estimation to gain extra advantage over population L by taking one route exclusively in one state, and splits in the other state. However, in class O, population L's expected cost difference between the two routes is not biased.

Then, this leverage is not possible, and population H has to split on both routes in both states unless λ^{H} is small (i.e. $\lambda^{\mathrm{H}} < \bar{\Lambda}$).

We demonstrate the ordering of the population costs in each classes of regimes with cost parameters in Table 5.2. In this example, the threshold parameter is $\widehat{p} = 0.375$.



(a) Class I regimes: $p = 0.8$ (b) Class II regimes: $p = 0.2$ (c) Class O regimes: $p = 0.375$

Figure 5-2: Equilibrium population costs in three classes of regimes with $\eta^{\mathrm{H}} = 0.8$.

5.2.2 Impact of perceived accuracy on traveler costs

We know from Theorem 1 that in class O regimes, η^{H} has no impact on the equilibrium strategies, and hence has no impact on the costs. The next proposition shows how population costs as in (5.10a) – (5.10b) change with η^{H} in class I and II regimes.

Proposition 3. *In regimes ϕ^1 and ϕ_1, if $\min\{\frac{\alpha_1^{\mathrm{n}}}{N}, \widehat{p}\} < p < \max\{\frac{\alpha_1^{\mathrm{n}}}{N}, \widehat{p}\}$, then $C^{\mathrm{H}*}$ increases in η^{H}. Otherwise, $C^{\mathrm{H}*}$ decreases in η^{H}. In regimes ϕ^2 and ϕ_2, $C^{\mathrm{H}*}$ increases in η^{H}.*

In regimes ϕ^1 (resp. ϕ_1), if $\beta_2 - \beta_1 + \left(\alpha_2 - (1-p)N - 2\rho(1-\eta^{\mathrm{H}})\right)\lambda^{\mathrm{H}}\mathrm{D} > 0$, then $C^{\mathrm{L}*}$ increases (resp. decreases) with η^{H}. Otherwise, $C^{\mathrm{L}*}$ decreases (resp. increases) with η^{H}. In regime ϕ^2, if $2N\mathrm{D}\Lambda^2 - K - \alpha_2\mathrm{D} - \alpha_1^{\mathrm{n}}\lambda^{\mathrm{H}}\mathrm{D} > 0$, then $C^{\mathrm{L}*}$ increases with η^{H}. Otherwise, $C^{\mathrm{L}*}$ decreases with η^{H}. In regime ϕ_2, $C^{\mathrm{L}*}$ decreases with η^{H}.

In regimes ϕ^3 – ϕ^4, and ϕ_3 – ϕ_4, both $C^{\mathrm{H}*}$ and $C^{\mathrm{L}*}$ do not change with η^{H}.

We now summarize the impact of perceived accuracy on population costs, and the intuition behind.

In regimes ϕ^1 and ϕ_1, the impact of η^{H} can be both positive and negative for both populations. From Theorem 1, we know that regardless of η^{H}, population H takes r_2 in state a and r_1 in state n. As η^{H} increases, more travelers (resp. fewer travelers) in population L take r_1 in regime ϕ^1 (resp. ϕ_1). Therefore, H faces lower cost in state a (resp. state n), and higher cost in state n (resp. state n). The monotonicity of cost with respect to η^{H} is determined by whether the cost increase in one state dominates the cost decrease in the other state or not. We find that if the probability of state a is relatively medium (i.e. $\min\{\frac{\alpha_1^{\mathrm{n}}}{N}, \widehat{p}\} < p < \max\{\frac{\alpha_1^{\mathrm{n}}}{N}, \widehat{p}\}$) then $C^{\mathrm{H}*}$ increases in η^{H} (i.e. $C^{\mathrm{H}*}$ is the minimum if L believes that H is also uninformed $\eta^{\mathrm{H}} = 0.5$), but not otherwise (i.e. $C^{\mathrm{H}*}$ is the minimum in the case with objective belief $\eta^{\mathrm{H}} = 1$).

For population L, the impact of η^{H} on the population cost is in opposite direction in regimes ϕ^1 and ϕ_1. As η^{H} increases, we can check that the total expected cost faced by travelers of population L who takes r_1 increases (resp. decreases), and that for travelers on r_2 decreases in regime ϕ^1 (resp. increases in regime ϕ_1). Therefore, how the average cost of travelers in L changes with η^{H} is determined by whether the cost increase on one route dominates the decrease on the other route or not.

We find that if $\beta_2 - \beta_1 + \left(\alpha_2 - (1-p)N - 2\rho(1-\eta^{\mathrm{H}})\right)\lambda^{\mathrm{H}}\mathrm{D} > 0$, then $C^{\mathrm{L}*}$ increases (resp. decreases) with η^{H} in regime ϕ^1 (resp. ϕ_1), and decreases (resp. increases) otherwise. Note that this condition is satisfied in both regimes if λ^{H} is close to zero. That is, when very few travelers are informed, if the probability of state a is high ($p > \widehat{p}$), then the cost of the uninformed travelers is the minimum if they believe that all other travelers are also uninformed ($\eta^{\mathrm{H}} = 0.5$). If the probability of state a is low ($p < \widehat{p}$), then the cost $C^{\mathrm{L}*}$ is the minimum in the case with objective belief.

In regimes ϕ^2 and ϕ_2, population H is always better off as η^{H} decreases. The population cost is the minimum if population L believes that H is also uninformed, i.e. $\eta^{\mathrm{H}} = 0.5$. This is because population H can take the less congested route exclusively in one of the two state by leveraging population L's perception bias, and H faces lower cost in that state as the perceived accuracy decreases.

On the other hand, more travelers in population L take r_1 (resp. r_2) as η^{H} increases. In regime ϕ^2 (resp. regime ϕ_2), this change makes no difference in state a

(resp. n) because the costs of both routes are the same. However, as η^{H} increases, travelers on r_1 (resp. r_2) face higher cost in state n, and travelers on r_2 (resp. r_1) face lower cost. Therefore, how $C^{\mathrm{L}*}$ changes with η^{H} depends on whether the cost increase on one route dominates the cost decrease on the other route in state a (resp. n).

We find that in regime ϕ_2, $C^{\mathrm{L}*}$ always decreases as the perceived accuracy increases. Therefore, $C^{\mathrm{L}*}$ is the minimum in the case with objective belief. However, in regime ϕ^2, $C^{\mathrm{L}*}$ increases if $2N\mathrm{D}\Lambda^2 - K - \alpha_2\mathrm{D} - \alpha_1^{\mathrm{n}}\lambda^{\mathrm{H}}\mathrm{D} > 0$. From (5.4a), we know that Λ^2 increases as η^{H} increases, and approaches to Λ^3 as η^{H} goes to 1. Additionally, $\lambda^{\mathrm{H}} \geq \Lambda^1$. Therefore,

$$2N\mathrm{D}\Lambda^2 - K - \alpha_2\mathrm{D} - \alpha_1^{\mathrm{n}}\lambda^{\mathrm{H}}\mathrm{D} \leq 2N\mathrm{D}\Lambda^3 - K - \alpha_2\mathrm{D} - \alpha_1^{\mathrm{n}}\Lambda^1\mathrm{D} = \beta_2 - \beta_1 - \alpha_1^{\mathrm{n}}\frac{K(\alpha_1^{\mathrm{a}} - \alpha_1^{\mathrm{n}})}{AN}.$$

If the slope difference on r_1 between state a and n is large, i.e. $\alpha_1^{\mathrm{a}} - \alpha_1^{\mathrm{n}} > \frac{(\beta_2-\beta_1)NA}{\alpha_1^{\mathrm{n}}K}$, then in ϕ^2, we always have $2N\mathrm{D}\Lambda^2 - K - \alpha_2\mathrm{D} - \alpha_1^{\mathrm{n}}\lambda^{\mathrm{H}}\mathrm{D} < 0$, i.e. $C^{\mathrm{L}*}$ decreases in η^{H} . That is, the cost of population L is the minimum in the case with objective belief. Otherwise, population L may face lower cost with subjective belief.

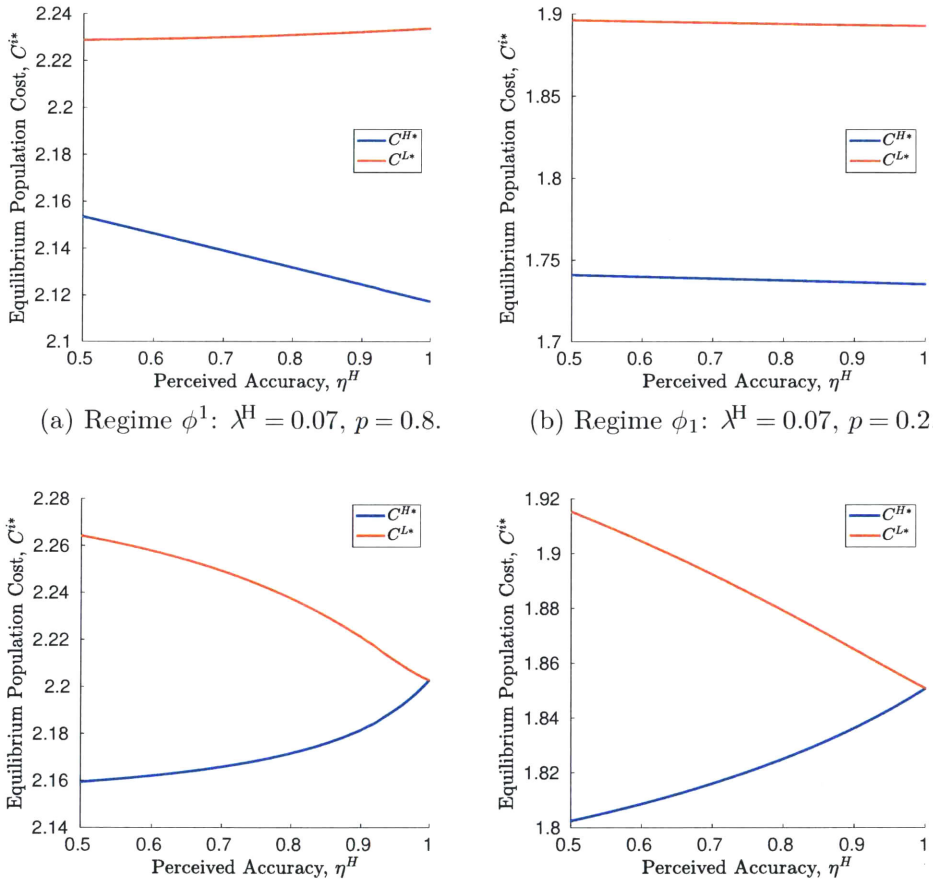
Finally, η^{H} has no impact on the population costs in regimes ϕ^3– ϕ^4 and ϕ_3 – ϕ_4. From Theorem 1, we know that in both classes of regimes, once the fraction of the informed population exceed a certain threshold (threshold Λ^2 in class I and threshold Λ_2 in class II), then population L exclusively takes one route (r_2 in class I and r_1 in class II). Therefore, the perceived accuracy does not change the strategy of population L, and hence does not change the route flow and costs.

We demonstrate the change of the equilibrium population costs with respect to η^{H} in regimes ϕ^1 - ϕ^2, and ϕ_1 - ϕ_2. The cost parameters are as in Table 5.2. The threshold probability is $\widehat{p} = 0.375$.

We can check that for $\lambda^{\mathrm{H}} = 0.07$, the condition $\beta_2 - \beta_1 + \left(\alpha_2 - (1-p)N - 2\rho(1-\eta^{\mathrm{H}})\right)\lambda^{\mathrm{H}}\mathrm{D} > 0$ is satisfied for both $p = 0.8$ in regime ϕ^1 and $p = 0.2$ in regime ϕ_1. Additionally, $p = 0.8 > \max\{\frac{\alpha_1^{\mathrm{n}}}{N}, \widehat{p}\}$, and $p = 0.2 < \min\{\frac{\alpha_1^{\mathrm{n}}}{N}, \widehat{p}\}$. Therefore, in regime ϕ^1, $C^{\mathrm{H}*}$ decreases with η^{H}, and $C^{\mathrm{L}*}$ increases (Figures 5-3a). In regime ϕ_1, both $C^{\mathrm{H}*}$ and $C^{\mathrm{L}*}$

decrease with η^{H} (Figure 5-3b).

Since our cost parameters satisfy the condition $\alpha_1^{\mathrm{a}} - \alpha_1^{\mathrm{n}} > \frac{(\beta_2 - \beta_1)NA}{\alpha_1^{\mathrm{n}} K}$, the equilibrium population cost $C^{\mathrm{H}*}$ increases, and $C^{\mathrm{L}*}$ decreases with η^{H} in both ϕ^2 (Figure 5-3c) and ϕ_2 (Figure 5-3d).



(a) Regime ϕ^1: $\lambda^{\mathrm{H}} = 0.07$, $p = 0.8$.

(b) Regime ϕ_1: $\lambda^{\mathrm{H}} = 0.07$, $p = 0.2$.

(c) Regime ϕ^2: $\lambda^{\mathrm{H}} = 0.4$, $p = 0.8$.

(d) Regime ϕ_2: $\lambda^{\mathrm{H}} = 0.4$, $p = 0.2$.

Figure 5-3: Impact of η^{H} on equilibrium population costs.

The following proposition shows how C^* changes with the perception η^{H} in class I and class II regimes.

Proposition 4. *In regime ϕ^1, if $\beta_2 - \beta_1 - 2\rho(1-\eta^{\mathrm{H}})\lambda^{\mathrm{H}}\mathrm{D} > 0$, then C^* increases in η^{H}. Otherwise, C^* decreases with η^{H}. In regimes ϕ^2, if $2N\mathrm{D}\Lambda^2 - K - \alpha_2\mathrm{D} > 0$, then C^* increases in η^{H}. Otherwise, C^* decreases in η^{H}.*

In regimes ϕ_1 – ϕ_2, C^* decreases in η^{H}.

In regimes ϕ^3 – ϕ^4, and ϕ_3 – ϕ_4, C^* does not change with η^{H}.

The impact of η^{H} on the average equilibrium cost can be computed by averaging the impact of η^{H} on the cost of each population with the weight of its size. Following Proposition 3, it is straightforward to see that η^{H} does not change the average cost in equilibrium in regimes ϕ^3 – ϕ^4, and ϕ_3 – ϕ_4.

Interestingly, if the probability of state a is high ($p > \widehat{p}$), then subjective belief with η^{H} that is slightly below 1 can result in a lower average equilibrium cost compared with the case of objective belief. Specifically, in regime ϕ^1,

$$\beta_2 - \beta_1 - 2\rho(1-\eta^{\mathrm{H}})\lambda^{\mathrm{H}}\mathrm{D} < \beta_2 - \beta_1 - 2\rho(1-\eta^{\mathrm{H}})\Lambda^1\mathrm{D},$$

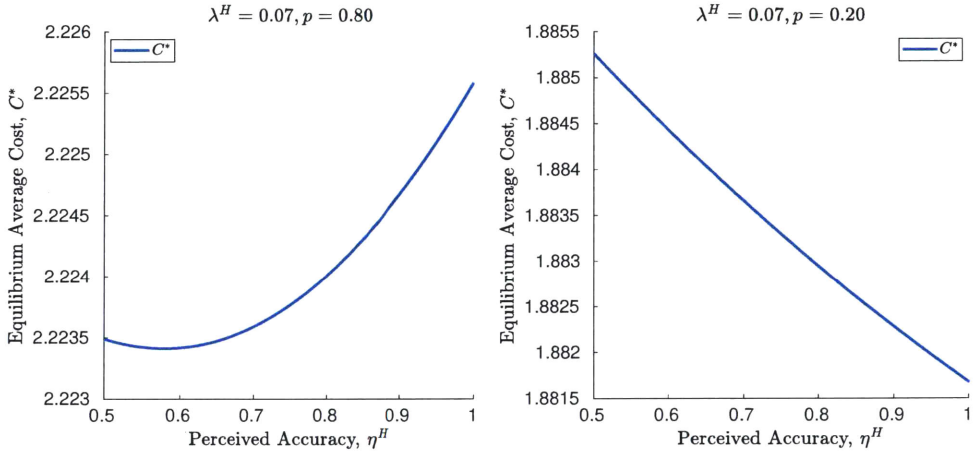
which increases in η^{H}, and equals to $\beta_2 - \beta_1 > 0$ when $\eta^{\mathrm{H}} = 1$. We can identify a threshold of η^{H} above which C^* increases in η^{H}. In regime ϕ^2, $2N\mathrm{D}\Lambda^2 - K - \alpha_2\mathrm{D}$ increases in η^{H}, and arrives at $\beta_2 - \beta_1 > 0$ when $\eta^{\mathrm{H}} = 1$. Hence, we can identify another threshold of η^{H} above which C^* increases in η^{H}. Therefore, in regimes ϕ^1 – ϕ^2, the average equilibrium cost with perceived accuracy η^{H} that is above the greater threshold identified in the two regimes but below 1 is lower than the average cost in the case with objective belief.

If the probability of state a is low ($p < \widehat{p}$), then the average equilibrium cost is the minimum in the case with objective belief.

We demonstrate the impact of η^{H} on the average equilibrium cost in Figure 5-4. We can see that in regimes ϕ^1 and ϕ^2 of class I, the perceived accuracy that minimizes the equilibrium average cost is below 1, see Figures 5-4a and 5-4c. In regimes ϕ_1 and ϕ_2 of class II, the average equilibrium cost is the minimum in the case with objective beliefs.

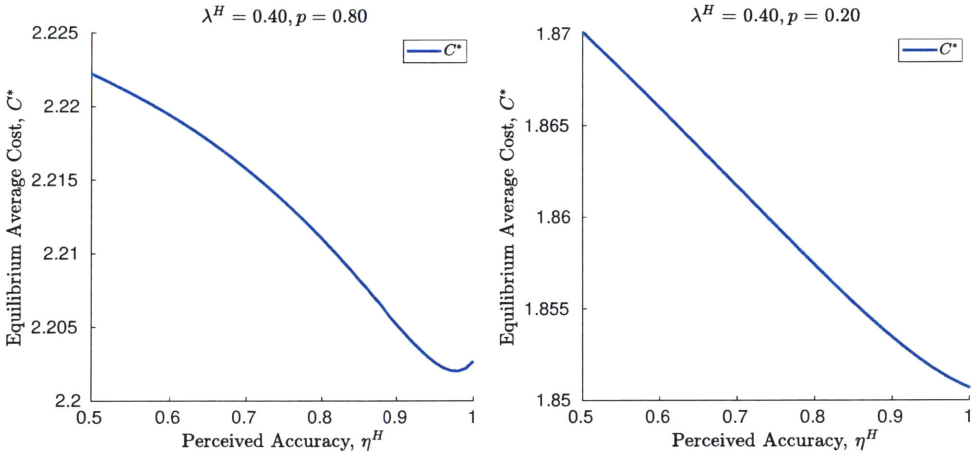
5.3 Solution Method for Subjective BCG

Whereas there exists a weighted potential function exists for the Objective BCG [104], no such potential function exists for the Subjective BCG. Thus, we must find an alternative approach for finding equilibrium solutions. In general, the BWE conditions



(a) Regime ϕ^1: $\lambda^{\mathrm{H}} = 0.07$, $p = 0.8$.

(b) Regime ϕ_1: $\lambda^{\mathrm{H}} = 0.07$, $p = 0.2$.



(c) Regime ϕ^2: $\lambda^{\mathrm{H}} = 0.4$, $p = 0.8$.

(d) Regime ϕ_2: $\lambda^{\mathrm{H}} = 0.4$, $p = 0.2$.

Figure 5-4: Impact of η^{H} on equilibrium average cost.

for a population game can be written as variational inequalities (VI) following [107]:

$$(q' - q^*)^\top F(q^*) \geq 0, \quad \forall q' \in Q \tag{VI}$$

where q^* is a BWE solution, and the function $F(q) = \left(\mathbb{E}_\mu[c_\mathrm{r}(q^*)|t^i]\right)_{t^i \in T^i, i \in I, \mathrm{r} \in R}$, where $\mathbb{E}_\mu[c_\mathrm{r}(q^*)|t^i]$ is expected cost of taking route r based on the belief of type t^i in equilibrium, see (4.11).

We note that on its own, (VI) does not provide a direct solution method; since Q is continuous, it is infeasible to enumerate through $\forall q \in Q$ and write an inequality for each one. However, it can be shown that it is sufficient to check only a finite subset $\widehat{Q} \subset Q$ in order to satisfy (VI). In fact, to check whether or not a feasible strategy profile is a BWE, we only need to check finite number of constraints. We start with defining the set of direction vectors $Z = \{z_\mathrm{r}^{t^i}\}_{t^i \in T^i, i \in I, \mathrm{r} \in R}$, where $z_\mathrm{r}^{t^i}$ is given as follows:

$$z_\mathrm{r}^{t^i} = \begin{bmatrix} 0 \\ \vdots \\ -1 \quad \text{— element corresponding to } q_\mathrm{r}^i(t^i) \\ \vdots \\ +1 \quad \text{— element corresponding to } q_{\mathrm{r}'}^i(t^i) \\ \vdots \\ 0 \end{bmatrix}, \tag{5.11}$$

where r' is the other route. That is, if a strategy profile q is perturbed in the direction of $z_\mathrm{r}^{t^i}$, then travelers in population i with type t^i shift demand from route r to the other route. There are in total $|R| \times |T^i| \times |I| = 2 \times 2 \times 2 = 8$ vectors in the set Z. However, not all vectors in Z are a feasible perturbation direction, if q is not an interior point of Q. We hence define the set $\bar{Z}_q$ as the set of feasible perturbation directions at q:

$$\bar{Z}_q = \{z_\mathrm{r}^{t^i} \in Z | \forall \mathrm{r} \in R, \forall t^i \in T^i, \text{ such that } q_\mathrm{r}^i(t^i) > 0\}. \tag{5.12}$$

The next proposition shows that it is sufficient to check a finite number of constraints associated with vectors in $\bar{Z}_q$.

Proposition 5. *A feasible strategy profile is a BWE if and only if it satisfies the following constraints:*

$$z^T F(q^*) \geq 0, \quad \forall z \in \bar{Z}_{q^*}. \tag{5.13}$$

Proof. Proof of Proposition 5 Clearly, if q^* is a BWE, i.e. q^* satisfies constraints in (VI), then q^* satisfies (5.13).

To prove the other direction, we first show that given any feasible strategy profile $q \in Q$, we can find a non-negative vector $\alpha = (\alpha_z)_{z \in \bar{Z}_{q^*}}$, where $\alpha_z \geq 0$ for all $z \in \bar{Z}_{q^*}$, such that

$$q - q^* = \sum_{z \in \overline{Z}_{q^*}} \alpha_z z. \tag{5.14}$$

We can check that the set of perturbations for r_1, $\{z_1^{t^i}\}_{t^i \in T^i, i \in I}$, is linearly independent. Additionally, the dimension of the subspace $\{q - q^* | \forall q \in Q\}$ is $\sum_{i \in I} |T^i|$. Hence, the set $\{z_1^{t^i}\}_{t^i \in T^i, i \in I}$ is a basis for the subspace, and for any $q \in Q$, we can find a unique vector $\beta = (\beta^{t^i})_{t^i \in T^i, i \in I}$ such that

$$q - q^* = \sum_{t^i \in T^i} \sum_{i \in I} \beta^{t^i} z_1^{t^i}. \tag{5.15}$$

For any $t^i \in T^i$ and $i \in I$, if $\beta^{t^i} \geq 0$, then we define $\alpha_1^{t^i} = \beta^{t^i}$, and $\alpha_2^{t^i} = 0$. Otherwise, we define $\alpha_1^{t^i} = 0$ and $\alpha_2^{t^i} = -\beta^{t^i}$. Hence, the elements in the vector $\alpha = (\alpha_{\mathrm{r}}^{t^i})$ are all non-negative, and (5.15) can be re-stated as follows:

$$q - q^* = \sum_{t^i \in T^i} \sum_{i \in I} \sum_{\mathrm{r} \in R} \alpha_{\mathrm{r}}^{t^i} z_{\mathrm{r}}^{t^i}. \tag{5.16}$$

For any $\mathrm{r} \in R$, $t^i \in T^i$, and $i \in I$ such that $q_{\mathrm{r}}^{i*}(t^i) = 0$, we know the vector $z_{\mathrm{r}}^{t^i} \notin \overline{Z}_{q^*}$, from (5.12). Since the element $q_{\mathrm{r}}^{i}(t^i) - q_{\mathrm{r}}^{i*}(t^i)$ must be non-negative, and β^{t^i} must

be non-positive, we know that in (5.16), $\alpha_{\mathrm{r}}^{t^i} = 0$. Therefore, by eliminating the term corresponding to $z_{\mathrm{r}}^{t^i}$ in (5.16), we show that (5.14) holds and α is a non-negative vector.

Finally, if $z \in \overline{Z}_{q^*}$ satisfy (5.13), then we obtain, from (5.14):

$$(q-q^*)^T F(q^*) = \sum_{z \in \overline{Z}_{q^*}} \alpha_z z^T F(q^*) \geq 0.$$

Hence, we can conclude that (5.13) is a sufficient condition for q^* to be a BWE. □

We note that the expression on the left-hand side of (5.13) is quadratic in q^*, and it is convex iff its Hessian matrix is positive semi-definite (PSD). It is straightforward to check that in the subjective case, the mixed derivatives are not equal, and thus, the Hessian is non-symmetric. Since it is necessary for a matrix to be symmetric to be PSD, it follows that the Hessian in the subjective case is not PSD, and that the set of solutions of (5.13) is non-convex.

We are able to find feasible BWE solutions by posing the problem as the following nonlinear program:

$$\begin{aligned} &\min_{q^*} 0 \\ &\text{subject to: } z^T F(q^*) \geq 0, \quad \forall z \in \bar{Z}_{q^*}. \end{aligned} \tag{5.17}$$

We apply sequential quadratic programming [108] to numerically solve (5.17) for more general configurations. As a heuristic, we use the Bayesian Wardrop Equilibrium solution of the Objective BCG for the initial guess. We were able to obtain numerical solutions for Subjective BCGs where both populations received information of different accuracies, and had different perceived accuracies of the other population's information.

5.4 Discussion

The Subjective BCG provides a game theoretic model of route choices in a disruption-prone network where travelers have heterogeneous information about the network state and different perceived beliefs of other travelers. This captures travelers' uncertainty about the information available to other travelers. We show that the Objective BCG with the conventional common prior is a special edge case of the Subjective BCG, and the solutions vary significantly depending on the misperception and subjectivity. The fact that population costs and average social costs can be affected by the perception of information suggests that public perception of information systems may also have indirect effects on the equilibrium route choices of travelers.

The model of subjectivity we use in this dissertation assumed that all members of a population had identical, discrete beliefs about the state (and other players' types). While this was convenient for formulating and examining effects of information, it is unrealistic that an entire population would have identical beliefs. As such, additional work is needed to extend the formulation to allow for a continuum of beliefs within each population of others' information. Furthermore, we considered only subjectivity in the accuracy of others' information, but it is also possible to consider subjectivity in other parameters: perception of likelihood of incident, size of each population, and other game parameters. This work is intended to be a demonstration of incorporating a boundedly-rational view of subjectivity into Bayesian Congestion Games, as well as an initial exploration of the resulting effects. Future work is necessary to develop richer, more comprehensive models which capture the many aspects of traffic information and their effects on route choices in game theoretic models.

Chapter 6

Conclusion

6.1 Summary

In summary, this dissertation addressed two questions related to disruptions in traffic networks. Part I addressed the question of identifying disruptions from existing traffic cameras. Toward this, it presented a novel, semantics-based approach to representing sequential image data using semantic labels in BoLW, and extracting *topics* and *topic signals* from the label data using the LDA topic model. These semantic representations represent the unstructured image data in a format that is both easily-interpretable by humans, as well as well-structured for statistical and signal processing analysis. The semantic representations offer a new approach to analyzing visual image data using off-the-shelf image labeling and object recognition tools.

Furthermore, we demonstrated applications of information retrieval and notable event detection using the BoLW and topic signal representations, and validated their performance against real-world data from our BFCC dataset. The topic signals represent latent visual phenomena, such as traffic and weather, as low-dimensional time series, which can be analyzed using signal processing techniques. We demonstrate that the topic signals aggregates information from several labels, and results in better performance for notable event detection, with less noise and better sensitivity to notable events than individual label signals.

Part II addressed the question of modeling the effects of information about traffic

disruptions on the route choices of travelers. In particular, it presented a Bayesian extension to the classical congestion game which incorporates populations of heterogeneously informed travelers making routing decisions in a network with stochastic cost functions. We presented two variants, the Objective and Subjective BCG. The Objective BCG adopts the common prior, which assumes that all travelers correctly know the accuracy of all other travelers' information. This is restrictive, as it cannot model situations where travelers are unaware of other information services. The Subjective BCG is a more general formulation which allows travelers to have a subjective "perceived accuracy" of other travelers' information.

We solve the Subjective BCG and compare it to the Objective BCG, demonstrating that the effect of subjective perceived accuracy has a significant effect on the equilibrium solution choices and costs, resulting in qualitatively distinct solutions between the two cases. We also find that the effects of subjective beliefs can be characterized succinctly by the value of a "misperception factor" parameter. Analysis of the solutions reveal that depending on the probability of incident, and the fraction of travelers with access to accurate information, it can be either beneficial or harmful for other travelers to know the true accuracy of your information about the network state. This can potentially inform decisions about whether to publicly advertise the accuracy of one's own information.

Furthermore, unlike the Objective BCG, the Subjective BCG does not admit a convex weighted potential function. We show that the solution set can be characterized by a finite set of variational inequalities, and is non-convex. This is able to be solved by an iterative, nonlinear numerical solver.

6.2 Practical Considerations and Future Directions

With regards to Part I, we emphasize the crux of our semantics-oriented analysis approach: analyzing semantic representations of image contents strongly resembles NLP problems. We provide the BoLW model as a foundational equivalent to the NLP BoW model. However, we acknowledge that many decades have passed since

the original BoW model, and there are now many more sophisticated models for representing semantic features in the NLP literature. In particular, concepts such as semantic word embeddings [109] can provide vector space representations of semantic features in fewer dimensions than BoLW.

Likewise, there exist more recent topic models than LDA, such as those that model conditional relationships between topics [110, 111]. Indeed, using these more sophisticated models could allow for better performance in the tasks presented in this dissertation, and is a promising avenue for future work. The intent of this dissertation is not to claim that these methods are the best approach for the tasks, but instead to establish a foundational proof-of-concept of using semantics-oriented analysis of image contents upon which others can build.

We intentionally used exclusively semantic features in order to explore the capabilities of semantic representations on their own. The low dimensionality of the semantic representations allow for convenient, quickly searchable indexing for information retrieval applications. However, in practice, we do not expect that purely-semantic representations will be ideal for most applications (aside from, perhaps, applications with strict privacy or bandwidth requirements, as semantic representations do provide some data de-identification and compression). Instead, we believe that semantic features will be complementary to visual features. Integrating BoLW semantic features to enhance existing BoVW computer vision applications to construct a multi-modal Bag-of-Features model should be a promising direction for future work.

Furthermore, we used a pretrained, off-the-shelf, general-purpose image labeling implementation to generate the semantic label features. This was done purposefully to focus on techniques downstream of the image labeling tasks. However, there is likely significant room for improvement if the image labeling implementation were trained to be domain-specific and tailored toward traffic applications. This represents a potential direction for future work. In particular, the ability of topic models to aggregate information from multiple labels and sources suggests that labels from general-purpose classifiers could be used to augment labels and other measurements from domain-specific classifiers and sensors.

We also note one of the most significant challenges encountered in the semantics-oriented analysis of image data. Our change and anomaly detection struggled with changes in camera angles and perspectives. Changes in perspective affect the distributions of image contents, and thus changed the distribution of labels and topics in the scene. This triggers change detection due to the shift in mean value, but could be worked around by reinitializing the change detection every time the camera angle changed. This also affects anomaly detection, because the reference data samples were no longer representative of the nominal conditions for that perspective. As such, all samples would get flagged as anomalous until the perspective returned to the original angle in the reference set. One possible workaround would be to have a separate reference set for each possible camera angle—though, this is only feasible if there are only a finite set of possible perspectives for each camera. Thus, additional work is required in order to account for the effects of camera angle changes. Such work could allow for the fusion of information from images from moving perspectives as well, such as vehicle-mounted cameras for autonomous driving or mapping applications.

With regards to Part II, our analysis of the Subjective BCG demonstrates that the common prior assumption is an edge case of a spectrum of possible beliefs. While the intent is not for the reader to completely reject the common prior assumption outright, but instead impart that the common prior should be carefully treated a *special* instance, rather than a conventional default. As such, the common prior assumption should be adopted *only if* there is reason to believe that players realistically have the knowledge implied by the assumption.

The model that was solved in Part II was a stylized instance, where only one population had access to perfectly accurate information. This was done intentionally to explore the effect of subjective information perception. However, there is potential for future work to explore the model in more general information environments: for example, where each population has access to information with different accuracies, and has different perceptions of others' information accuracies. Furthermore, the beliefs of each population were assumed to be identical. In practice, we would expect there to be a continuous distribution of beliefs about both the state and other players'

information. This can be accounted for by relaxing the player population sets and type sets to be continuous instead of discrete, and replacing the respective sums with integrals. However, additional work is needed to determine which formulations are tractable. In addition, extensions to additional routes and number of populations may be informative. The analytic derivation of the solutions presented in this part is possible for extensions with additional players and routes, but was tedious to derive manually even for the simplified, stylized instance presented in Part II. Thus, analyses of more complex configurations may have to be approach numerically, or with symbolic computation.

This work also only considered subjectivity in the beliefs of others' information accuracy. However, in practice, travelers may have subjective beliefs about other game parameters, such as the probability of incidents, their own information's accuracy, and which players have access to which information. Our work demonstrates how to incorporate subjectivity of game parameters in the game formulation and how it impacts the solutions. The model could be extended to incorporate subjectivity in multiple parameters by expanding the "perceived accuracy" variable to be a "perceived parameters" vector, which accounts for each population's beliefs of the values of each of the game parameters. The Objective BCG serves as an important reference point in this expanded subjective formulation as the instance of the game where all of the perceived parameters align with the true parameter values of Nature. Additional work is necessary to determine which formulations are tractable, and how each of the additional dimensions of subjectivity affect the solution.

Finally, there is potential work in studying the impact of subjective perception in learning dynamics, where travelers repeatedly commute in the network and aggregate information of the unknown state based on the realized costs. In such setting, since subjective perception impacts the travelers' route choices and the realized costs, subjective perception may effect the information that is available to travelers in each stage, and whether or not the true parameter values can be learned eventually. Such models could also be used to investigate system behavior over time, such as modeling market adoption of traffic information services, and modeling the response to changes

in the network parameters such as construction projects.

Together, the work presented in the two parts of this dissertation contribute toward addressing disruptions in transportation networks. Improving the ability to detect anomalies in various phenomena from the widely-deployed networks of traffic cameras in a human-parseable way allows operators at management agencies to better recognize and respond to disruptions. Extending the capabilities of congestion game models to incorporate heterogeneous information and subjective beliefs about information allows researchers to model not only the direct effects of information, but also the comparably impactful aspect of how information is presented and perceived. This work seeks to serve and model uniquely human considerations in transportation through semantics and subjectivity.

Bibliography

[1] S. Kuciemba and K. Swindler, "Transportation Management Center Video Recording and Archiving Best General Practices," U.S. Department of Transportation Federal Highway Administration, Tech. Rep., 2016.

[2] R. Marois and J. Ivanoff, "Capacity limits of information processing in the brain," *Trends in Cognitive Sciences*, vol. 9, no. 6, pp. 296–305, Jun. 2005.

[3] V. Kastrinaki, M. Zervakis, and K. Kalaitzakis, "A survey of video processing techniques for traffic applications," *Image and Vision Computing*, vol. 21, no. 4, pp. 359–381, 2003.

[4] N. Buch, S. A. Velastin, and J. Orwell, "A review of computer vision techniques for the analysis of urban traffic," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 3, pp. 920–939, 2011.

[5] C.-N. E. Anagnostopoulos, I. E. Anagnostopoulos, I. D. Psoroulas, V. Loumos, and E. Kayafas, "License plate recognition from still images and video sequences: A survey," *IEEE Transactions on intelligent transportation systems*, vol. 9, no. 3, pp. 377–391, 2008.

[6] Y. A. LeCun, Y. Bengio, and G. E. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[7] R. Livni, S. Shalev-Shwartz, and O. Shamir, "On the Computational Efficiency of Training Neural Networks," *NIPS*, pp. 1–15, 2014.

[8] R. Girshick, I. Radosavovic, G. Gkioxari, P. Dollár, and K. He, "Detectron," 2018. [Online]. Available: https://github.com/facebookresearch/detectron

[9] Tensorflow, "ResNet in TensorFlow," 2018. [Online]. Available: https://github.com/tensorflow/models/tree/master/official/resnet

[10] Y. Jia, "Caffe Model Zoo," *Caffe Model Zoo*, 2015. [Online]. Available: http://caffe.berkeleyvision.org/model_zoo.html

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," Jun. 2015. [Online]. Available: http://arxiv.org/abs/1506.02640

[12] Google, "google-cloud vision 963db69 documentation." [Online]. Available: https://perma.cc/6BN7-E247

[13] Microsoft, "Computer Vision API - Image Processing | Microsoft Azure," 2018. [Online]. Available: https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/

[14] Amazon Web Services, "Amazon Rekognition – Video and Image - AWS," 2018. [Online]. Available: https://aws.amazon.com/rekognition/

[15] T. Pamula, "Road Traffic Conditions Classification Based on Multilevel Filtering of Image Content Using Convolutional Neural Networks," *IEEE Intelligent Transportation Systems Magazine*, vol. 10, no. 3, pp. 11–21, 2018.

[16] Z. Zhang, Q. He, J. Gao, and M. Ni, "A deep learning approach for detecting traffic accidents from social media data," *Transportation Research Part C: Emerging Technologies*, vol. 86, pp. 580–596, Jan. 2018.

[17] Z. Luo, P.-M. Jodoin, S.-Z. Su, S.-Z. Li, and H. Larochelle, "Traffic Analytics With Low-Frame-Rate Videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 4, pp. 878–891, Apr. 2018.

[18] M. Yamada, T. Suzuki, T. Kanamori, and M. Sugiyama, "Relative Density-Ratio Estimation for Robust Distribution Comparison," Tech. Rep. 5, 2013.

[19] S. Zhang, G. Wu, J. P. Costeira, and J. M. Moura, "Understanding traffic density from Large-ScaleWeb camera data," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Jan, 2017, pp. 4264–4273.

[20] A. P. Shah, J.-B. Lamare, T. Nguyen-Anh, and A. G. Hauptmann, "CADP: A novel dataset for CCTV traffic camera based accident analysis," *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–9, 2018.

[21] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[22] Google, "Google Cloud Vision API enters Beta, open to all to try!" 2016. [Online]. Available: https://perma.cc/D6LF-NQ4X

[23] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, vol. II. IEEE, 2005, pp. 524–531.

[24] C. D. Manning, P. Ragahvan, and H. Schutze, "An Introduction to Information Retrieval," *Information Retrieval*, no. c, pp. 1–18, 2009.

[25] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[26] T. L. Griffiths and M. Steyvers, "Finding scientific topics," *Proceedings of the National Academy of Sciences*, vol. 101, no. Supplement 1, pp. 5228–5235, 2004.

[27] F. Wang, T. Xu, T. Tang, M. Zhou, and H. Wang, "Bilevel Feature Extraction-Based Text Mining for Fault Diagnosis of Railway Systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 1, pp. 49–58, Jan. 2017.

[28] K. D. Kuhn, "Using structural topic modeling to identify latent topics and trends in aviation incident reports," *Transportation Research Part C: Emerging Technologies*, vol. 87, pp. 105–122, Feb. 2018.

[29] S. Roller, S. Schulte, and I. Walde, "A Multimodal LDA Model Integrating Textual, Cognitive and Visual Modalities," *EMNLP*, pp. 1146–1157, 2013. [Online]. Available: http://stephenroller.com/research/

[30] S. Workman, R. Souvenir, and N. Jacobs, "Wide-area image geolocalization with aerial reference imagery," in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2015 Inter, Oct. 2015, pp. 3961–3969.

[31] V. J. Hodge and J. Austin, "A survey of outlier detection methodologies," Tech. Rep. 2, 2004.

[32] A. P. Tewkesbury, A. J. Comber, N. J. Tate, A. Lamb, and P. F. Fisher, "A critical synthesis of remotely sensed optical image change detection techniques," *Remote Sensing of Environment*, vol. 160, pp. 1–14, 2015.

[33] T. Suzuki, S. Shirakabe, Y. Miyashita, A. Nakamura, Y. Satoh, and H. Kataoka, "Semantic Change Detection with Hypermaps," Apr. 2016. [Online]. Available: http://arxiv.org/abs/1604.07513

[34] M. Sugiyama, S. Nakajima, H. Kashima, P. Buenau, and M. Kawanabe, "Direct Importance Estimation with Model Selection and Its Application to Covariate Shift Adaptation," *NIPS*, pp. 1–8, 2007.

[35] T. Kanamori, S. Hido, and M. Sugiyama, "A Least-squares Approach to Direct Importance Estimation," *Journal of Machine Learning Research*, vol. 10, pp. 1391–1445, 2009.

[36] S. Hido, Y. Tsuboi, H. Kashima, M. Sugiyama, and T. Kanamori, "Statistical Outlier Detection Using Direct Density Ratio Estimation," Tech. Rep. 2, 2011.

[37] S. Liu, M. Yamada, N. Collier, and M. Sugiyama, "Change-point detection in time-series data by relative density-ratio estimation," *Neural Networks*, vol. 43, pp. 72–83, Jul. 2013.

[38] W. H. Sandholm, "Large population potential games," *J. Economic Theory*, vol. 144, no. 4, pp. 1710–1725, jul 2009.

[39] R. W. Rosenthal, "A class of games possessing pure-strategy Nash equilibria," *International Journal of Game Theory*, vol. 2, no. 1, pp. 65–67, dec 1973.

[40] T. Roughgarden, "Routing games," *Algorithmic game theory*, vol. 18, 2007.

[41] ——, *Selfish routing and the price of anarchy*. MIT press Cambridge, 2005, vol. 174.

[42] D. Monderer and L. S. Shapley, "Potential games," *Games and economic behavior*, vol. 14, no. 1, pp. 124–143, 1996.

[43] W. H. Sandholm, "Potential games with continuous player sets," *Journal of Economic theory*, vol. 97, no. 1, pp. 81–108, 2001.

[44] J. G. Wardrop, "Road paper. some theoretical aspects of road traffic research." *Proceedings of the institution of civil engineers*, vol. 1, no. 3, pp. 325–362, 1952.

[45] R. B. Dial, "A probabilistic multipath traffic assignment model which obviates path enumeration," *Transportation Research /UK/*, vol. 5 N2, Jun. 1971.

[46] C. F. Daganzo and Y. Sheffi, "On stochastic models of traffic assignment," *Transportation science*, vol. 11, no. 3, pp. 253–274, 1977.

[47] M. G. H. Bell, "Stochastic user equilibrium assignment in networks with queues," *Transportation Research Part B: Methodological*, vol. 29, no. 2, pp. 125–137, Apr. 1995.

[48] W. H. K. Lam, Z. Y. Gao, K. S. Chan, and H. Yang, "A stochastic user equilibrium assignment model for congested transit networks," *Transportation Research Part B: Methodological*, vol. 33, no. 5, pp. 351–368, Jun. 1999.

[49] R. Cominetti, F. Facchinei, and J. B. Lasserre, "Wardrop and stochastic user equilibrium," in *Modern Optimization Modelling Techniques*. Springer, 2012, pp. 213–220.

[50] D. Fudenberg and J. Tirole, "Game theory. 1991," *Cambridge, Massachusetts*, vol. 393, 1991.

[51] D. Acemoglu, A. Makhdoumi, A. Malekian, and A. Ozdaglar, "Informational braess' paradox: The effect of information on traffic congestion," *Operations Research*, 2018.

[52] I. Milchtaich, "Congestion Games with Player-Specific Payoff Functions," *Games and Economic Behavior*, vol. 13, no. 1, pp. 111–124, 1996.

[53] T. Roughgarden, "The price of anarchy in games of incomplete information," in *Proceedings of the 13th ACM Conference on Electronic Commerce*. ACM, 2012, pp. 862–879.

[54] S. Morris, "The common prior assumption in economic theory," *Economics & Philosophy*, vol. 11, no. 2, pp. 227–253, 1995.

[55] R. Arnott, A. De Palma, and R. Lindsey, "Does providing information to drivers reduce traffic congestion?" *Transportation Research Part A: General*, vol. 25, no. 5, pp. 309–318, 1991.

[56] Z. Khan and S. Amin, "Bottle neck model with heterogeneous information," 2018, forthcoming.

[57] M. Ben-Akiva, A. de Palma, and I. Kaysi, "The impact of predictive information on guidance efficiency: An analytical approach," in *Advanced methods in transportation analysis*. Springer, 1996, pp. 413–432.

[58] H. S. Mahmassani and R. Jayakrishnan, "System performance and user response under real-time information in a congested traffic corridor," *Transportation Research Part A: General*, vol. 25, no. 5, pp. 293–307, 1991.

[59] M. Ben-Akiva, A. De Palma, and K. Isam, "Dynamic network models and driver information systems," *Transportation Research Part A: General*, vol. 25, no. 5, pp. 251–266, 1991.

[60] T. Roughgarden and É. Tardos, "Bounding the inefficiency of equilibria in nonatomic congestion games," *Games and Economic Behavior*, vol. 47, no. 2, pp. 389–403, 2004.

[61] E. Koutsoupias and C. Papadimitriou, "Worst-Case Equilibria," in *STACS 99*, ser. Lecture Notes in Computer Science, C. Meinel and S. Tison, Eds. Springer Berlin Heidelberg, 1999, vol. 1563, pp. 404–413.

[62] J. R. Correa, A. S. Schulz, and N. E. Stier-Moses, "Fast, fair, and efficient flows in networks," *Operations Research*, vol. 55, no. 2, pp. 215–225, 2007.

[63] D. Acemoglu and A. Ozdaglar, "Competition and efficiency in congested markets," *Mathematics of Operations Research*, vol. 32, no. 1, pp. 1–31, 2007.

[64] E. Nikolova and N. E. Stier-Moses, "A mean-risk model for the traffic assignment problem with stochastic travel times," *Operations Research*, vol. 62, no. 2, pp. 366–382, 2014.

[65] R. Heumen, B. Peleg, S. Tijs, and P. Borm, "Axiomatic characterizations of solutions for bayesian games," *Theory and Decision*, vol. 40, no. 2, pp. 103–129, 1996.

[66] G. Facchini, F. van Megen, P. Borm, and S. Tijs, "Congestion models and weighted bayesian potential games," *Theory and Decision*, vol. 42, no. 2, pp. 193–206, 1997.

[67] M. Mavronicolas, I. Milchtaich, B. Monien, and K. Tiemann, "Congestion games with player-specific constants," in *International Symposium on Mathematical Foundations of Computer Science*. Springer, 2007, pp. 633–644.

[68] D. Blackwell *et al.*, "Comparison of experiments," in *Proceedings of the second Berkeley symposium on mathematical statistics and probability*. The Regents of the University of California, 1951.

[69] J. Hirshleifer, "The private and social value of information and the reward to inventive activity," *The American economic review*, vol. 61, no. 4, pp. 561–574, 1971.

[70] R. Haenfler, "Social value of public information," *The American Economic Review*, vol. 92, no. 5, pp. 1521–1534, 2002.

[71] A. Neyman, "The positive value of information," *Games and Economic Behavior*, vol. 3, no. 3, pp. 350–355, 1991.

[72] M. Goldstein *et al.*, "Subjective bayesian analysis: principles and practice," *Bayesian analysis*, vol. 1, no. 3, pp. 403–420, 2006.

[73] J. C. Harsanyi, "Games with Incomplete Information Played by Bayesian Players, I-III Part I. The Basic Model," *Management science*, vol. 14, no. 3, pp. 159–182, 1967.

[74] R. J. Aumann, "Agreeing to Disagree," in *The Annals of Statistics*. Institute of Mathematical Statistics, 1976, vol. 4, pp. 1236–1239.

[75] J. Liu, A. Weinert, and S. Amin, "Semantic topic analysis of traffic camera images," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 568–574.

[76] ——, "Semantic Analysis of Traffic Camera Data: Topic Signal Extraction and Anomalous Event Detection," *Submitted to Transactions of Intelligent Transportation Systems*, 2019.

[77] M. J. Menne, I. Durre, B. Korzeniewski, S. McNeal, K. Thomas, X. Yin, S. Anthony, R. Ray, R. S. Vose, B. E. Gleason, and Others, "Global historical climatology network-daily (GHCN-Daily), Version 3," pp. subset: December 2017–January 2018, 2012. [Online]. Available: https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00861

[78] T. Andersen and E. Sweeney, "Parking ban being lifted at 5 p.m. in Boston - The Boston Globe," 2018. [Online]. Available: https://perma.cc/Q47F-ZWFT

[79] MassLive. Christmas Eve / Christmas 2017: What's open, what's closed. [Online]. Available: https://perma.cc/CH9R-CFB5

[80] ——. (2018) New Year's Day 2018: What's open, what's closed. [Online]. Available: https://perma.cc/3LU7-6QX4

[81] ——. (2017) Veterans Day 2017: What's open? What's closed? Hours for post office, Walmart, Target, supermarkets, CVS, Rite Aid, Walgreens, more | masslive.com. [Online]. Available: https://perma.cc/V288-EMZ7

[82] ——. (2018) Martin Luther King Jr. Day 2018: What's open and what's closed | masslive.com. [Online]. Available: https://perma.cc/U2M4-YRNT

[83] Mass511, "Cameras | Traffic Cameras - Mass511." [Online]. Available: https://mass511.com/list/cctv

[84] M. Zhang and Z. Zhou, "A Review on Multi-Label Learning Algorithms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 8, pp. 1819–1837, Aug. 2014.

[85] Z. Lin, G. Ding, M. Hu, and J. Wang, "Multi-label Classification via Feature-aware Implicit Label Space Encoding," *Proceedings of The 31st International Conference on Machine Learning*, vol. 32, pp. 325–333, Jan. 2014.

[86] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017.

[87] Z. S. Harris, "Distributional structure," *Word*, vol. 10, no. 2-3, pp. 146–162, 1954.

[88] J. Yang, Y.-G. Jiang, A. G. Hauptmann, and C.-W. Ngo, "Evaluating bag-of-visual-words representations in scene classification," in *Proceedings of the international workshop on Workshop on multimedia information retrieval - MIR '07*. New York, New York, USA: ACM Press, 2007, p. 197.

[89] L. Van Der Maaten and G. Hinton, "Visualizing Data using t-SNE," Tech. Rep., 2008.

[90] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *Journal of the American Society for Information Science*, vol. 41, no. 6, pp. 391–407, Sep. 1990.

[91] T. L. Griffiths, M. Steyvers, and J. B. Tenenbaum, "Topics in semantic representation." *Psychological review*, vol. 114, no. 2, p. 211, 2007.

[92] M. Hoffman, D. Blei, and F. Bach, "Online learning for latent dirichlet allocation," in *Advances in Neural Information Processing Systems*, 2010, pp. 856–864.

[93] W. Zhao, J. J. Chen, R. Perkins, Z. Liu, W. Ge, Y. Ding, and W. Zou, "A heuristic approach to determine an appropriate number of topics in topic modeling." *BMC bioinformatics*, vol. 16 Suppl 1, no. Suppl 13, p. S8, 2015.

[94] @CityOfBoston, "SNOW ON THE WAY: A Snow Emergency & Parking Ban will go into effect for #Boston at 7 a.m. Thursday. For information on parking & other winter resources, go to http://boston.gov/winter," 2018. [Online]. Available: https://twitter.com/CityOfBoston/status/948656142976212994

[95] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning*, ser. Springer Series in Statistics, M Jordan, J. Kleinberg, and B. Scholkopf, Eds. Springer, 2001, vol. 27, no. 2.

[96] S. Salvador and P. Chan, "Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms," in *16th IEEE International Conference on Tools with Artificial Intelligence*. Boca Raton, FL, USA: IEEE Comput. Soc, 2004, pp. 576–584.

[97] M. Lavielle, "Using penalized contrasts for the change-point problem," *Signal Processing*, vol. 85, no. 8, pp. 1501–1510, Aug. 2005.

[98] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 131–142, 1966.

[99] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[100] K. Pearson, "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 50, no. 302, pp. 157–175, 1900.

[101] R. W. Rosenthal, "Correlated equilibria in some classes of two-person games," *International Journal of Game Theory*, vol. 3, no. 3, pp. 119–128, 1974.

[102] M. Wu, J. Liu, and S. Amin, "Informational aspects in a class of bayesian congestion games," in *American Control Conference (ACC), 2017*. IEEE, 2017, pp. 3650–3657.

[103] J. Liu, M. Wu, and S. Amin, "Aspects of subjective perception of travelers' information in Bayesian Congestion Games," *Submitted to Economics of Transportation*, 2019.

[104] M. Wu, S. Amin, and A. E. Ozdaglar, "Value of information systems in routing games," *arXiv preprint arXiv:1808.10590*, 2018.

[105] G. Brown, "Iterative solutions of games by fictitious play." *In T. Koopmans (Ed.), Activity Analysis of Production and Allocation, Volume 13 of Cowles Commission Monograph, New York: Wiley*, pp. 374–376, 1951.

[106] A. Garcia, D. Reaume, and R. Smith, "Fictitious play for finding system optimal routings in dynamic traffic networks," *Transportation Research Part B*, vol. 34, pp. 147–156, 2000.

[107] W. H. Sandholm, "Population games and deterministic evolutionary dynamics," in *Handbook of game theory with economic applications.* Elsevier, 2015, vol. 4, pp. 703–778.

[108] P. T. Boggs and J. W. Tolle, "Sequential quadratic programming," *Acta numerica*, vol. 4, pp. 1–51, 1995.

[109] Y. Bengio, H. Schwenk, J. S. Senécal, F. Morin, and J. L. Gauvain, "Neural probabilistic language models," *Studies in Fuzziness and Soft Computing*, vol. 194, pp. 137–186, 2006.

[110] D. M. Blei, J. D. Lafferty *et al.*, "A correlated topic model of science," *The Annals of Applied Statistics*, vol. 1, no. 1, pp. 17–35, 2007.

[111] M. Roberts, "The structural topic model and applied social science," *NIPS 2013 Workshop on Topic Models*, pp. 2–5, 2013.