INTEGRATING SOCIAL AND ENVIRONMENTAL PERSPECTIVES IN BUILDING DESIGN AND OPERATION

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF
CIVIL AND ENVIRONMENTAL ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Andrew James Sonta
November 2020

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Rishee Jain, Primary Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Martin Fischer

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

Michael Lepech

Approved for the Stanford University Committee on Graduate Studies.

Stacey F. Bent, Vice Provost for Graduate Education

This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.

Abstract

In commercial buildings, for every \$1 spent on energy (as measured through utilities), \$10 are spent on space (the real estate), and \$100 are spent on people (their salaries, wages, and benefits). This economics is well known by managers of organizations and those in the real estate industry, but it is often overlooked by researchers and practitioners in the energy efficiency domain. It is also well known that buildings have an outsized impact on the environment—an impact that is expected to continue growing. But because people are so much more expensive than energy, any rational view toward sustainable commercial building design and operation must balance the two perspectives of organizational performance and energy performance. Generally, researchers have considered these two perspectives separately. An opportunity to integrate them is to enhance our understanding of the relationship between buildings and occupants, as occupant behavior fundamentally drives building operation, and the dynamics of interactions among individuals drive the success of organizations. **The overarching motivation for this dissertation is to work toward integrating social and environmental perspectives in sustainable building design and operation.**

A key opportunity for enabling this integration is the emergence of sensing and data within buildings. Data from ambient sensors, when analyzed properly, can offer valuable insight into building operation—including how building occupants use spaces in buildings. Measuring occupant dynamics through data has direct applications for real-time building management, but it also enables new types of research that consider the social behavior of building occupants and the implications of occupant behavior on building design. These are the key research areas that are investigated in this dissertation.

The first area I investigate is how sensing can be used to improve our methods for inferring occupant behavioral dynamics. In this work, I use *plug load energy sensors*, which, when installed at the individual desk level, offer insight into occupants' use of individual spaces. I introduce a novel data mining approach that integrates data-driven analysis with explicit engineering knowledge on office workstation systems. This approach clusters the time series sensor data into abstracted *states* of occupant activities. The result is a sensing strategy that is real-time and unsupervised, and because it outputs *states* of activity, it offers additional information beyond simple presence and absence.

Once occupant behavioral dynamics can be inferred in real time, analysis of social behavior becomes possible at scale. The second research question I investigate is whether occupant socio-organizational networks can be inferred from these activity patterns. I introduce a new method, the Interaction Model, that models *opportunities for social interaction* among building occupants and uses them as building blocks to form the overall socio-organizational network. When compared to network data obtained through a survey, I find that this inferred network correlates with the survey network to a statistically significant degree. Modeling building occupants as a network enables additional methods for analyzing building operation.

Occupant dynamics are exceedingly complex in part because they involve temporal, spatial, and social dimensions. The third area of research in this dissertation involves the creation of a novel framework based on discrete signal processing that enables identification of *out-of-sync* occupant behavior. I show that such instances of behavior can be used to create realignment strategies that improve the cohesiveness of building operation and occupant dynamics.

The fourth area of investigation concerns the direct design of building layouts. Once occupant dynamics can be measured at scale, a natural question is whether the layout of the building can be improved such that controllable building systems (e.g., lighting, heating, and cooling systems) can better respond to occupant space use and therefore save energy. I show that optimization can be used to create new building layout designs that enable substantial reduction of building energy consumption.

Overall, the research in this dissertation contributes toward a design and operation framework for the built environment in which environmental goals (e.g., energy efficiency) are integrated with social goals (e.g., organizational performance). In the context of commercial buildings, design and operation that explicitly consider knowledge of occupant behavior and organizational dynamics can lead to improved building operation and organizational performance. The results and methods I introduce have theoretical and practical implications that will enable a more environmentally and socially sustainable built environment.

Acknowledgements

Attempting to integrate ideas across disciplines, as this research does, is to require the support of many individuals—too many to thank fully, though I will risk a brief set of acknowledgements all the same. My primary advisor, Rishee Jain, has supported this research endeavor from the start. His thoughtfulness, encouragement, feedback, and support during my doctoral studies have been foundational to any success I have achieved. There is no better advisor.

My reading committee members, Michael Lepech and Martin Fischer, who welcomed me to Stanford, have supported me with expertise and guidance throughout my graduate-school journey. John Taylor, my additional oral exam committee member, has been a genuine mentor both from afar, and (when luck had it) when we were in the same location. My oral exam chair, Amir Goldberg, sent me down fruitful paths as I, an engineer, attempted to enter the world of social science.

I have been fortunate through my graduate studies to be surrounded by the colleagues, mentors, and friends that are the members of the Urban Informatics Lab. Jonathan Roth was there from the very beginning, before the UIL even had a physical location, and has supported my work with inquisitiveness and friendship. Zheng Yang, also there at the beginning, always had the perfect paper or reference when I was stuck on something. Alex Nutkiewicz, Rohan Aras, Ranjitha Shivaram, Marco Miotti, Abigail Andrews, and Kopal Nihar have all been lab-mates, colleagues, and friends of the highest quality. The research in this dissertation was not mine alone—it was only possible with the research acumen of my collaborators and mentees: Perry Simmons, Amy Kouch, David Showunmi, and Thomas Dougherty. Their leadership on their respective aspects of this work has been top notch; what shortcomings exist are mine alone. Thomas: I am on the edge of my proverbial seat with excitement to see how you transform and extend the ideas in this work. I have also been lucky to collaborate with researchers beyond Stanford: Rimas Gulbinas, José Moura, and John Taylor, whose expertise has furthered this research.

I am indebted to the institutions that have supported this research in myriad tangible ways. Each chapter has benefited from an organization that has been generous enough to either share data or allow me to outfit their building with sensors. They are unnamed in this dissertation, and, unfortunately, I suppose that means I cannot thank them enough. At Stanford, Moira Zbella in the Office of Sustainability has been an unstoppable resource for all my data collection efforts. The Department of Civil and Environmental Engineering and the Sustainable Design and Construction

Program only exist because of the people who make them run: Jill Filice (if I learned one thing at Stanford, it is that Jill knows all), Brenda Sampson, Emilia Alex, and many others. Financially, this work has been supported by Stanford University through a William and Martha Campbell Stanford Graduate Fellowship, the Center for Integrated Facility Engineering, and the National Science Foundation.

My family and friends at Stanford and beyond: I am beyond grateful to you. I would like to thank: My extended Sustainable Design and Construction cohort, for making Stanford feel like home. Friends near and far, many of whom either shared a home with me as roommates or offered a couch when I was traveling, for keeping at least one of my feet on the ground. 2039, for Looz et al. The Potrero and Berkeley crews, for making the (increasingly virtual) commute worth it. My sister, Elisabeth, for being the best big sister and a role model (and a *real* doctor).

My partner, Montana Marshall: you are an unwavering source of support, compassion, and encouragement; my adventure buddy and my home base; my biggest cheerleader; my inspiration.

This dissertation is dedicated to my parents, Mary and Jack Sonta, who encouraged me to dream big, and were always there whenever anything seemed too big.

For my parents

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Much of our human experience is shaped by buildings. This is especially true in the United States, where we spend about 90% of our time indoors (Klepeis et al. 2001). Buildings, therefore, strongly impact the subjective experiences of their occupants, in both good and bad ways. In return, people, in their presence, actions, and patterns of behavior, have strongly impact the operation of buildings. This two-way interaction between our built and social systems is complex and poorly understood. If it were better understood, our buildings would operate more efficiently, and people would have fewer complaints about their indoor experiences. Unfortunately, inefficiency and poor experiences of the built environment are common and well-documented (Norford et al. 1994; Keeling et al. 2012).

The problem of inefficiency in building operation is a direct form of energy waste. Amid growing concerns of global climate change, buildings have been identified as critical opportunities for reducing global energy demand. This environmental concern is at the heart of much research on the built environment today (Lucon et al. 2014). Unfortunately, the outcomes of this research often come at the expense of social systems. As an extreme example, a perfectly energy-efficient building would not operate at all, leaving the indoor thermal and visual environment uninhabitable. More commonly, the engineering solutions that improve operating efficiency of buildings simply do not regard expected impacts on the experiences of occupants, outside basic guidelines for occupant comfort (Krukar et al. 2016). Our attempts to improve our human or social experiences of the built environment, on the other hand, do not typically concern the issue of energy. This is largely driven by economics: for organizations that occupy buildings, the cost of people (i.e., salaries, wages, and benefits) tends to be around 100 times higher than the cost of energy (i.e., utilities) (*Revenue and expense data, University of California* n.d.).

I have identified two key goals for buildings: to support our social systems and to reduce impact on the environment. This dual responsibility for buildings, to address both environmental and social goals, forms the overarching motivation for this dissertation (Figure 1.1). As we spend more time indoors and as cities continue to grow, it is increasingly becoming the case that our built environment intermediates our social world and our natural environment: that our social world is contained within what is built and our natural environment outside it. As urban environments continue to grow, with more than 2/3 of the world's population expected to live in cities by 2050 (United Nations 2014), we will construct and retrofit buildings to hold our homes and workplaces, with these built spaces dictating the bulk of our existence. The work I present here focuses on this dual responsibility in the context of commercial buildings, which house complex social systems—often in the form of organizations—and which are responsible for 20% of global energy demand (U.S. Energy Information Administration 2016a). As a result, the specific research motivations for this dissertation are to improve commercial building energy efficiency through operation of energy-intensive building systems (the environmental goal) and to better model and understand the structure of organizations to help improve their success (the social goal).

Figure 1.1: Responsibility of the built environment to meet both social and environmental objectives.

Each goal I have identified for the built environment holds a strong research tradition, which has formed the foundation for the research in this dissertation. Research at the intersection of architecture and organizational behavior has shown that the design of buildings can strongly influence characteristics of organizational success. For example, spatial configuration has been connected to collaboration (Kabo et al. 2015), creativity (Sailer 2011), and other aspects of productivity (Peponis et al. 2007). This past research suggests that incorporating aspects of organizational structure in spatial design can promote components of overall organizational success. Simultaneously, a wealth of research in building design and engineering has shown the many ways in which buildings can be improved for energy efficiency. All aspects of buildings, from site planning, to building form and materiality, to operation of lighting systems, have been rigorously investigated for the purposes of reducing energy consumption. Recently, researchers have identified the behavior of building occupants as one of the key reasons for unexpected inefficiencies in building operation (Feng et al. 2015). Occupants often use spaces in ways that were not modeled or intended, they override building systems, and they generally behave in ways that are difficult to model (Yan et al. 2015). Analyzing and understanding occupant behavior is a primary objective for improving building energy efficiency.

This focus on the occupant as a means to address energy efficiency opens a new possibility to directly consider our social systems in building energy research. People are at the heart of both organizations and building operation, thereby offering a shared reference point. If we can understand how people and organizations use the spaces of buildings, we can work toward an ability to design and manage these spaces in a way that promotes our environmental and social goals simultaneously. An opportunity to enhance our understanding of human behavior in buildings is the growing prevalence of data. Data created within buildings, often from ambient sensors, can be used to paint a spatially and temporally granular picture of building operation (Melfi et al. 2011). Some of this data can be used to shed light on human use of built spaces. Ultimately, the research

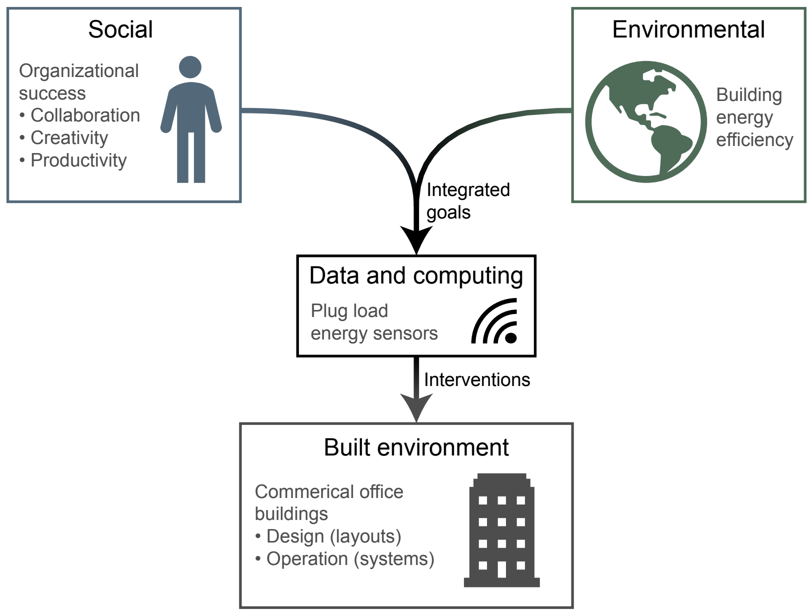

Figure 1.2: Overarching thesis framework.

encompassed in this dissertation aims to use such data to create an understanding of the social and environmental consequences of building design and management.

Figure 1.2 shows the overarching framework driving the research in this dissertation. The key theoretical motivation is the integration of social and environmental goals for our built environment. Sensing data provides a new path forward to analyze each objective and create this integration. Ultimately, the analysis of data through the lens of social and environmental goals enables new ways of intervening in our built world. In the chapters that follow, I apply this framework to commercial buildings, where the primary social goal is the success of organizations, and the primary environmental goal is energy efficiency. I deploy a specific sensing strategy—plug load energy sensors installed at the desk level—to understand how building occupants use the spaces of commercial buildings. The analysis of the data produced by these sensors, through the lens of the building occupant, creates new ways to design and manage commercial building spaces for organizational success and energy efficiency.

1.1 Background

In this section, I briefly summarize the key areas of previous research that this dissertation builds upon: the importance of understanding occupant behavior and new strategies for doing so, the importance of understanding occupant social and organizational relationships, how these behaviors and networks relate to spatial design of buildings, and how building design can be improved using the lens of occupant dynamics.

1.1.1 The importance of understanding occupant behavior

Building performance is highly complex and dependent on many factors. One factor that has begun to receive more attention is occupant behavior, especially as it impacts energy efficiency. Research focused on commercial office spaces has identified two major components of total building energy consumption: baseline energy consumption and human-driven energy consumption—energy use that depends entirely on the behavior of building occupants (Taherian et al. 2010). Research has shown that occupants can be a large source of uncertainty in expected building energy consumption (Norford et al. 1994). The impact and uncertainty are largely driven by occupants' use of building spaces and interaction with building systems (Jia et al. 2017).

In addition to occupants' impact on building energy, the ways in which occupants use the spaces of a building can have a large influence on the architectural design process. The process of space planning generally involves the use of heuristics on how occupants are expected to use designed spaces (Dzeng et al. 2015). Recent work has underscored the notion that it is difficult to quantify and optimize the function of spaces due to a lack of information about how occupants utilize spaces designed for them. Dzeng et al. (2015) found that function space assignment optimizations that are based on user activities can increase the prescribed function objectives significantly.

Addressing the energy impacts of occupant behavior as well as improving models of space utilization depends on a solid understanding of how occupants utilize spaces in existing buildings. Only by monitoring and understanding the dynamics of occupant dynamics in existing buildings can we hope to imagine how occupants will behave in the new buildings we design ore renovate (Tomé et al. 2015).

1.1.2 Occupant sensing in buildings

The importance of understanding building occupant behavior creates a need for tools that enable occupant sensing. Researchers have proposed a variety of sensors and sensing strategies that provide varying levels of information granularity. Melfi et al. (2011) discuss the need for granularity along three dimensions of occupancy: space, time, and information about the occupant (Figure 1.3). The least granular information about the occupant is whether or not the space analyzed is occupied at all. Increasing granularity involves discovering richer information beyond simple presence or absence: initially, *how many* occupants are in the space, and ultimately the *activities* of these occupants. This highly granular information about occupant activities can be interpreted in multiple ways, depending on the purpose of the study. For some purposes, it may be useful to know whether occupants are

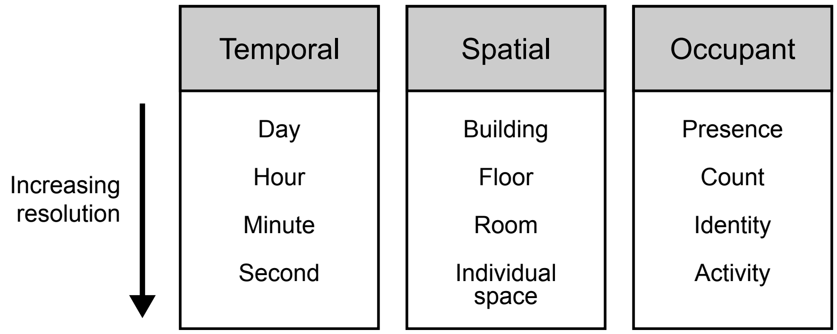

Figure 1.3: Dimensions of occupant sensing, adapted from Melfi et al. (2011).

interacting with building systems. For others, it may be useful to know whether they are using their workstations or interacting with other occupants in the building.

Spatial and temporal resolution can typically be reconciled through the sensor hardware. For example, increased frequency of interval reporting can improve temporal granularity, and increased density of sensors within the building can increase spatial granularity, within certain limits. Increasing the granularity of information about the occupant is a challenging process and often involves building models to analyze sensor data. Highly ambient sensors, such as infrared, carbon dioxide, and acoustic sensors, are limited in their ability to detect granular occupant information beyond occupancy count (Nguyen & Aiello 2012). On the other hand, information-rich sensors, such as video cameras, can reveal much when the data is processed, but come with significant privacy concerns (Jayarajah et al. 2016). Recently, researchers have noted the potential for plug load energy sensors (particularly in office buildings) to sense occupancy at highly granular resolutions across all three time, space, and occupant dimensions (Zhao et al. 2014). As discussed in detail in Chapter 2, these sensors reveal detailed information about occupant space use while preserving aspects of anonymity.

1.1.3 Occupant networks

The analysis of occupant behavior enables the study of further questions about what this behavior suggests. One such paradigm for human behavioral analysis is the social network: a mathematical representation of ties among individuals. In the context of buildings, this network has often been referred to as the *occupant network*.

The structure of occupant networks has been extensively studied in the context of building energy efficiency. Researchers have found that modeling occupant relational ties can be instrumental in analyzing energy-efficient occupant behavior. In particular, the structure of social networks can affect the efficacy of campaigns that encourage occupants to adopt more energy-efficient behaviors (Anderson et al. 2014; Azar & Menassa 2014; Jain et al. 2013).

Beyond energy efficiency, modeling building occupants with a network has a long tradition in organizational behavior research. It is well-known that simple organizational charts showing an organization's hierarchy lack a description of the subtleties in human interaction that fundamentally drive work (Krackhardt & Hanson 1993). True organizational structure is highly complex, and knowing this structure can enable more effective management of workspaces (Peponis et al. 2007).

While the benefits of modeling the occupant network are well known, a challenge remains as to how best learn its structure. Collection of social network data typically involves the use of surveys or interviews, which are possible for small studies but quickly become intractable in large study settings, such as large buildings or campuses of buildings. When it comes to commercial building management, a *data-driven* method for inferring the occupant network would enable social and organizational analysis at much larger scales—in the contexts of both energy efficiency and organizational behavior. This is discussed in detail in Chapter 3.

1.1.4 Spatial configuration and work outcomes

Knowledge of occupant behavioral dynamics and socio-organizational structure naturally relates to the design of commercial building spaces. Architectural researchers have argued that the ordering of space is really about the ordering of relationships among people (Hillier & Hanson 1984). As discussed above, these spatial relationships have been shown to correlate with different components of organizational outcomes. For example, researchers have shown that stronger spatial relationships between individuals correlate with more frequent and more successful collaboration (Kabo et al. 2015). As a result of this relationship between spatial and social systems, there remains a pressing need to able to analyze the dynamics of building occupants in the context of these systems. This is discussed in greater detail in Chapter 4.

1.1.5 Building design and energy efficiency

As the previous subsections have discussed, analyzing occupancy patterns with sensor data can be useful for operation of building systems as well as the study of the organization vis-à-vis building design. It is also important to recognize that the design of buildings has a large impact on building energy consumption. Energy-efficient building design is a highly researched area and has largely focused on early design decisions such as building orientation, materiality, and fenestration (Basbagill et al. 2014). However, one important aspect of building design that has not been investigated deeply in the context of energy efficiency is the spatial configuration of occupants.

In commercial buildings, the most energy-intensive building systems are those that provide heating, cooling, ventilation, and lighting (Lucon et al. 2014). These are generally shared resources: they create comfortable thermal and visual conditions for spaces that are often occupied by more than one individual. Understanding the patterns of each occupant's utilization of their space enables the analysis of the efficiency of the building layout. Therefore, the question remains as to whether building layouts can be improved to reduce the amount of time that building systems are providing these shared resources. For example, if occupants with similar schedules are grouped within the same lighting zones, that lighting zone can tailor its operation to match those schedules (D'Oca et al. 2017). If the occupant behavioral patterns are inferred for each occupant in a building, it may be possible to optimize the building layout to reduce overall energy consumption, as discussed in detail in Chapter 5. To further integrate social and environmental goals, these design decisions can also be informed by organizational network constraints.

1.2 Structure of dissertation

The research presented in this dissertation leverages ambient sensing to describe and analyze human use of the built environment and the resulting impact on building design. The underlying sensing strategy throughout the dissertation involves plug load energy sensors installed at the individual desk level in commercial office buildings, which enables analysis of individual occupant behavior. Each chapter includes a case study of a real office building, with study sizes ranging from 7 to 165 occupants. The following chapters seek to address four primary research objectives, described below. Each chapter was written as a standalone journal article—published or submitted—along with the co-authors noted at the start of the chapter.

Chapter 2: Inferring building occupant activities with ambient sensing data

This chapter addresses the well-known challenge of understanding how building occupants use space in commercial buildings. It highlights three key gaps in modern occupant sensing strategies:

1. The need for an unsupervised approach, to enable rapid extensibility to new settings
2. The importance of high spatial and temporal granularity in analyzing occupant behavior
3. The ability to report occupant activity patterns in real time, as information that can be useful for building managers

The chapter introduces a novel sensing methodology based on applying a variational Bayesian inference strategy with a Gaussian mixture model to plug load energy sensor data. It seeks to address the key research gaps by integrating this data-driven approach with explicit engineering knowledge on occupant dynamics.

Chapter 3: Inferring socio-organizational network structure of building occupants

This chapter builds upon the work presented in Chapter 2 by seeking to infer the structure of organizations in commercial buildings. Once individual occupant activity patterns can be characterized, following the research in Chapter 2, what do these patterns reveal about the social nature of occupants' behavior? Learning this network has been shown to be helpful in addressing both energy-related and organizational aspects of building management.

Chapter 4: Analyzing the complexity of occupant behavior

This chapter focuses on analyzing occupant behavior in the context of their spatial and social networks. Due to the multiple dimensions that can be attributed to occupant dynamics (i.e., spatial, temporal, and social), a nuanced understanding of occupant behavior is difficult. By embedding our expectations for occupant behavior in a network, and analyzing behavior continuously over time, we are able to determine when occupant behavior is *out-of-sync* with our expectations.

Chapter 5: Optimizing building layouts to reduce energy consumption

This chapter introduces a modeling framework for optimizing layouts of existing buildings based on occupant behavioral dynamics. It first considers the relationship between energy consumption and occupant space use, examining the hypothesis that increased diversity in occupant schedules within a particular building zone correlates with increased energy consumption. Once this relationship is established, the chapter introduces an optimization framework for reducing this diversity in building lighting zones by clustering occupants with similar schedules.

Chapter 6: Conclusion

This chapter summarizes the theoretical and practical contributions from Chapters 2–5 and discusses areas for future research.

Chapter 2

Understanding building occupant activities at scale

An integrated knowledge-based and data-driven approach

This chapter is adapted from the following paper: Sonta, A. J., Simmons, P. E., and Jain, R. K. (2018). "Understanding building occupant activities at scale: An integrated knowledge-based and data-driven approach." *Advanced Engineering Informatics*, 37, 1–13.

Abstract

Buildings are our homes and our workplaces. They directly affect our well-being, and they impact the natural global environment primarily through the energy they consume. Understanding the behavior of occupants in buildings has vital implications for improving the energy efficiency of building systems and for providing knowledge to designers about how occupants will utilize the spaces they create. However, current methods for inferring building occupant activity patterns are limited in two primary areas: First, they lack adaptability to new spaces and scalability to larger spaces due to the time and cost intensity of collecting ground truth data for training the embedded algorithms. Second, they do not incorporate explicit knowledge about occupant dynamics in their implementation, limiting their ability to uncover deep insights about activity patterns in the data. In this paper, we develop a methodology for classifying occupant activity patterns from plug load sensor data at the desk level. Our method makes use of a common unsupervised learning algorithm—the Gaussian mixture model—and, in addition, it incorporates explicit knowledge about occupant presence and absence in order to preserve adaptability and effectiveness. We validate our method using a pilot study in an academic office building and demonstrate its potential for scalability through a case study of an open-office building in San Francisco, CA. Our method offers key insights into spatially and temporally granular occupancy states and space utilization that could not otherwise be obtained.

2.1 Introduction

Buildings are integral to our daily lives. People spend an estimated 87% of their time indoors (Klepeis et al. 2001), and researchers have shown that buildings directly affect our well-being (Keeling et al. 2012). Moreover, buildings worldwide account for over 19% of energy-related CO_2 emissions and 51% of global energy consumption (Lucon et al. 2014), making them an integral part of our sustainable energy future. Fundamentally, buildings consume this energy to provide their occupants with services, including thermal and visual comfort, access to water, and power for electronic devices. As a result, understanding the relationship between buildings and their occupants is central to designing buildings that enhance occupant well-being, improve service delivery, and reduce energy usage.

We define occupant dynamics as the complex interactions between buildings and humans, encompassing occupant presence, occupant behavior (i.e., the specific actions that occupants take in buildings, such as working at a workstation, taking a break, or even interacting with lighting or heating, ventilation, and air conditioning (HVAC) controls), occupant activity states (i.e., abstracted and categorized information about occupant behavior), and the impact occupant behavior has on building operations. These dynamics are challenging to understand due to the increasing complexity of our building systems and the socio-technical complexities of occupant behavior. Even gaining a clear picture of the spatial and temporal activity patterns of occupants within a building is a non-trivial task (W. Shen et al. 2017). While new types of sensors have facilitated more data-driven approaches to understanding occupant-building dynamics, they suffer from a few key limitations. Sensors designed to directly detect occupancy often mischaracterize the spaces they are sensing due to the complexity associated with various building spaces (Gunay et al. 2016). New statistical and data mining techniques that have been proposed to infer occupancy patterns from emerging high-fidelity data streams such as light levels (Yang et al. 2012), energy use (Jin et al. 2014), sound (Khan et al. 2014), and video (Erickson et al. 2009) typically require a significant amount of ground truth training data that is cumbersome and often cost prohibitive to collect, thereby limiting their applicability and feasibility at scales beyond small pilot studies. Conversely, knowledge-based approaches to understanding occupant dynamics in buildings (e.g., surveys, on-site engineering audits) can yield insights on occupant dynamics (Andersen et al. 2009; Ingle et al. 2014) but suffer from common reliability and scalability issues associated with indirect collection instruments (Gunay et al. 2013).

In various aspects of building design, construction, and management involving human activity, researchers have shown that combining expert knowledge about buildings with automated computing techniques can vastly improve the effectiveness of the embedded methods. In the context of augmented reality within buildings, researchers have shown that integrating explicit engineering knowledge about building layout and operator movement into the automated augmented reality framework can improve the accuracy of the overall system (Neges et al. 2017). In construction management, the process of extracting meaningful information about the activities of construction workers from raw cellphone data can be enhanced by incorporating explicit engineering knowledge about the necessary levels of detail required for improving the effectiveness of construction activity simulations (Akhavian & Behzadan 2015). These studies and others like them emphasize the point that automated methods can be made more accurate and effective by integrating knowledge about the specific domain in the design of the overall methodology.

In this paper, we present a new methodology that integrates knowledge-based and data-driven approaches to understanding occupant activities in buildings with the goal of informing enhanced building design and energy efficient operations. Our method infers activity states for individual occupants using time-series data from low-cost, off-the-shelf plug load sensors. It incorporates explicit domain knowledge about how occupant activities impact plug load data into a common unsupervised learning algorithm—the Gaussian mixture model—to characterize the data into abstracted levels of activity. We design our method to be able to automatically analyze the highly variable data associated with occupant presence separately from the less variable data associated with occupant absence. This design decision in our method allows it to more deeply characterize the data while maintaining adaptability to new spaces, potential for scalability to larger spaces, and high accuracy. We validate and demonstrate that our method is able to determine individual occupancy states with a high-level of accuracy on a small control study, and we demonstrate the merits and applicability of our approach on a case study of a real 47-person open office in San Francisco, CA, USA.

2.2 Background

Building designers and managers are increasingly utilizing sensors and the data they collect to make decisions about how buildings are designed, built, and operated (Weng & Agarwal 2012). These sensors measure properties such as air temperature and humidity, lighting levels, sound, movement, and plug load energy use (Doukas et al. 2007; Singhvi et al. 2005; Li et al. 2010; Guo et al. 2010; Weng et al. 2011). Each of these types of sensors produces time-series data that provides information about the changing state of the building. In many cases, data produced within a building can be utilized to make decisions that can improve the energy efficiency of that building: for example, a lighting sensor may provide feedback to lighting controls that can dim the overhead lighting if the building is receiving enough light from outdoors. In others, data can be used to understand characteristics of existing buildings so that the design of future buildings can be improved: for example, data describing existing building occupancy can be linked with predictive energy models to increase the accuracy of energy models (Menezes et al. 2012).

This explosion of data has created an opportunity to provide new knowledge to engineers, designers, and building managers. In particular, previously unavailable information about the state of occupancy in buildings—the presence or absence of occupants as well as their activities—can be useful both for efficient building control of existing buildings and for improved space planning of future buildings (Roetzel et al. 2014). Along with other data streams specific to each building system, the detection of occupant activities has been shown to be significant in addressing all forms of energy use in buildings, from lighting control (Galasiu et al. 2007; Roisin et al. 2008; Williams et al. 2012) to HVAC control (Balaji et al. 2013; Dong et al. 2011). In addition, as knowledge about space use becomes more widely available to designers, the integration of design heuristics with occupancy models will be integral to designing spaces that better suit the needs of occupants (Roetzel 2015). In this section, we discuss the state of data-driven decision making in buildings for energy efficient building operations and improved building design, as well as the importance of occupancy and the state of the art for detecting occupant presence and occupant behavior in buildings. We elucidate the need for a robust, adaptable method for determining the activity states of occupants in buildings.

2.2.1 Data-driven & occupant-driven energy efficiency

Over recent years, the analysis of building energy data with statistical and data mining techniques has been shown to be helpful in improving energy efficient management of building systems. Within buildings, researchers have worked toward achieving a condition in which building systems—such as lighting, heating, and cooling—are provided only as much as they are needed, and only where and when they are needed. Matching these building systems with occupancy information has been shown to lead to significant energy savings (Singhvi et al. 2005; Li et al. 2010). Recently in commercial buildings, energy use data collected through power strips installed at the individual outlet level have been used for multiple approaches to save energy in buildings: to show that energy is wasted due to inefficient occupant behavior, such as leaving lights or other systems on during non-occupied hours (Masoso & Grobler 2010); to calibrate and improve the accuracy of building energy models in conjunction with other building data sources (Raftery et al. 2011); and to describe the behavior of occupants and improve schedule modeling in buildings (Zhao et al. 2014).

Many studies have noted the high impact occupant presence and behavior has on building energy use (Bonte et al. 2014; Norford et al. 1994; Taherian et al. 2010). Jia et al. (2017) has noted that occupant behavior (as distinct from occupancy) relates to more than just the presence or absence of occupants in buildings—that is, the activities of occupants within the building have a large impact on building energy performance. However, this human element, which is responsible for much of building energy use, is often difficult to characterize. One reason is because it is multidimensional, requiring a fundamental understanding of spatial, temporal, and social dimensions of occupant behavior (Sonta et al. 2017). Understanding each of these dimensions and reconciling their effects on occupant behavior is critical to gaining a broad understanding of occupant behavior and its impact on building energy use. Furthermore, the structure and type of the social network of occupants has been shown to be highly influential when it comes to how occupants behave and adapt to information in buildings (Anderson et al. 2014; Xu et al. 2014). Researchers have shown that providing the right information to occupants can lead to changes in behavior that reduce buildings' energy consumption (Staats et al. 2000, 2004; Staddon et al. 2016; Khosrowpour et al. 2016; Jain et al. 2012). Due to the energy-consumption impact, complexity, and ever-changing nature of occupant dynamics in buildings, there remains a pressing need to better understand them.

2.2.2 Occupancy data & space utilization

While whole-building data and occupancy data have typically been studied in the context of energy efficient management of existing buildings, they also have the potential to be tremendously useful in providing knowledge to designers in the early stages of building design. Previous research has utilized model-based optimization in the design of buildings (Mourshed et al. 2011), and more specifically, in the planning of space layouts in buildings (Jo & Gero 1998; Dzeng et al. 2015; Suter et al. 2014). Recent work has conceptualized models that utilize computing in the assessment of the functional properties of designed spaces (Bhatt et al. 2012). Specifically, analyzing designs for their ability to perform their function—for example, the ability for a proposed office space to promote a productive work environment—depends on knowledge from empirically based methods (e.g., surveys) (Bhatt & Schultz 2014).

Architects have traditionally used personal perceptions of how occupants will use the spaces they design in their planning process. Formalized integration of human-centered knowledge into the building design process has previously been focused on perceptions of space (Schultz et al. 2017) and heuristics for improved layouts (Bhatt et al. 2010), among others. More recent work has underscored the notion that it is difficult to quantify and optimize the function of spaces due to a lack of information about how occupants utilize spaces designed for them. Dzeng et al. (2015) found that function space assignment optimizations that are based on user activities can increase the prescribed function objectives significantly (e.g., improving overall space use by optimizing prescribed building assignments in a remodeling effort). However, these methods typically use occupant activity simulation models that are built on occupant activity data obtained through onerous methods such as defining heuristics from previous spaces and predetermined schedules (Dzeng et al. 2015) or from specialized occupant movement sensors (Dzeng et al. 2014). With the potential to accurately and granularly detect occupant activities in existing buildings from more ubiquitous sensors, new design-knowledge integration approaches will have greater opportunity to incorporate empirically grounded occupant activity patterns into new design heuristics.

As engineers and designers continue developing tools to aid in the design of buildings that more appropriately meet the needs of their occupants, analysis of the utilization of spaces has become increasingly important. Space-use analysis helps designers determine how appropriately the spaces within a building are serving their occupants. Spaces that are properly utilized fulfill their design intentions by having a certain level of occupancy at predetermined times and by not inhibiting occupants from performing predetermined activities. Spaces that are not properly utilized can either be underutilized (in which case they are inefficient in their use of space), or they can be too crowded (in which case they inhibit occupants from performing the activities they were meant to be able to perform), with proper utilization rates depending on the nature of the space being analyzed (Vischer 2008; Kim et al. 2013). Recently, researchers have proposed frameworks that can be helpful for architects working on space-utilization in the programming phase of their design process, but those frameworks depend on a detailed understanding of how occupants use the spaces designed for them (Kim et al. 2013). There remains significant opportunity to analyze the activities of occupants in existing spaces for greater understanding of occupant dynamics in planned spaces (Akbas et al. 2007). Furthermore, carefully representing the information gained from these analyses can provide useful knowledge to key decision-makers such as building designers or managers (Dijkstra & Timmermans 2002).

Further improvement of the accuracy of models that help designers understand how occupants will use the spaces they plan depends on a solid understanding of how occupants utilize spaces in existing buildings. Only by monitoring and understanding the dynamics of occupant dynamics in existing buildings can we hope to imagine how occupants will behave in the new buildings we design (Tomé et al. 2015).

2.2.3 Detecting occupant presence and activities in buildings

Because occupant behavior is so important to the energy use and space-use planning of buildings, there is a need for better tools to detect, model, and understand levels of occupancy and the activities of occupants. Melfi et al. (2011) discusses the need for understanding occupancy at a high level of granularity in terms of temporal resolution, spatial resolution, and resolution of occupancy (activities versus presence/absence). Specifically, as the level of resolution increases, more information can be gained from the sensors, elucidating the need for sensors that can determine the activities of occupants at the spatial resolution of individual workstations and the temporal resolution of minutes.

Recent work has utilized sensors or combinations of sensors—including infrared (Meyn et al. 2009), video (Trivedi et al. 2000), and acoustic (Kleine-Cosack et al. 2010)—to estimate the occupancy of rooms in buildings. These studies have shown that intelligently controlling systems such as lighting through use of occupancy sensors can save significant amounts of energy in buildings. Other recent work has used computer vision algorithms to characterize the movement of occupants in building spaces, noting that understanding the activities of building occupants leads toward a better understanding of how spaces can be designed for improved spatial efficiency (i.e., more properly utilized spaces) and better user experiences (Tomé et al. 2015; Krukar et al. 2016).

More recent work has utilized plug load energy data collected at the desk level as an additional input for algorithms that estimate the true occupancy levels of buildings (Amayri et al. 2016; Arora et al. 2015; Milenkovic & Amft 2013). Zhao et al. (2014) has shown that plug load data of computers and task lights at the desk level can be utilized to determine occupants' activities rather than just the level of occupancy. Due to the fact that plug load sensors are relatively inexpensive and often already installed in commercial office buildings for investigations into plug load management (Poll & Teubert 2012), they can be considered a low-cost alternative to sensors designed specifically for occupancy detection, such as infrared sensors. Moreover, many sensors that are designed specifically for occupant presence detection—such as infrared, acoustic, and CO2 sensors—require large time lags up to 60 minutes for high accuracy, while plug load sensors have been shown to be useful in determining occupant presence at time scales on the order of 5-15 minutes (Norford et al. 1994; Xu et al. 2014).

The analysis of plug load data for the detection of occupant activities typically involves data mining techniques and classification algorithms such as decision trees (Amayri et al. 2016; Arora et al. 2015) and hidden Markov models (Milenkovic & Amft 2013). These classification techniques are used to map the collected plug load energy use data to levels of occupancy, or in more sophisticated algorithms, to the types of activities occupants perform in buildings, such as working at a workstation. This recent work using plug load sensors has high value in advancing our methodologies for determining occupant activities, however, it requires training the classification models on ground truth data that is often onerous and cost-prohibitive to collect. While such previous work demonstrates how energy use data can be utilized to gain an understanding of occupant activities, it is limited in its ability to naturally adapt to quick changes in the appliances used at the desk or to additional workers in the space. Scaling to large buildings with many individuals would require large, coordinated efforts to collect ground truth data for training individual models. As a result, it is necessary to have a method that is robust and that can easily adapt to individual consumption patterns in order to glean a deeper understanding of occupant dynamics within a building. For this reason, this paper aims to contribute an effective methodology for detecting individual occupant activity patterns across a building. By combining knowledge-based and data-driven approaches, we are able to uncover occupant activity patterns from plug load energy use data and utilize such insights to inform energy efficient operational and space utilization strategies in a commercial building.

2.3 Methodology

Our occupant activity state classification method maps occupants' energy use, collected through plug load energy use sensors over a 5–20-minute period, to activity states. Based on major ideas from Section 2, our method is designed with a few key motivations in mind. The method should be able to adapt to different situations, such as when occupants in different buildings or floorplans tend to have different types of appliances at their workstations. For design of the algorithm, this adaptability would mean that parameters would not need to be tuned and set before the method is applied. In addition, the method should not require the use of training data in order to infer activity states, as we believe that a cost-effective and adaptable method should be able to be applied to new situations (such as new buildings or floorplans) without requiring onerous data collection procedures. In order to accomplish these tasks, our method is designed to combine a common unsupervised learning approach—the Gaussian mixture model—with explicit engineering knowledge about the typical structure of plug load energy signatures.

Various algorithms have been applied to time-series data—like that captured through plug load sensors—for the purposes of uncovering the underlying clusters in the data. In the context of analyzing plug load data, previous work has used supervised learning algorithms that train models on ground truth data and utilize these models to predict the correct classification of new data. In particular, the naïve Bayes and support vector machine algorithms have been shown to be successful in determining whether or not occupant workstations are occupied (Zhao et al. 2014). However, these supervised learning methods require the collection of ground truth data for training. Because we are interested in designing a method that does not require training on ground truth data, we looked toward unsupervised clustering algorithms that recognize patterns in the data and produce clusters with similar characteristics. Previous work has utilized the k-means clustering algorithm to disaggregate occupancy presence data into typical patterns of occupancy schedules (D'Oca et al. 2018). The k-means algorithm avoids the need to train underlying models on ground truth data, but it requires setting the number of clusters—"k"—before running the model. Therefore, the user must either know ahead of time how many clusters are present in the data, or tune the parameter by running the algorithm with various values of "k" and picking the model based on some goodness-of-fit test. To avoid the need to train on ground truth data and/or tune model parameters, we chose to build our method around variational Bayesian inference.

Our method requires collecting continuous time series energy use data from each occupant's workstation using a plug load sensor at a time granularity of 5-20 minutes, as this time scale adequately captures changes in occupant activities (Zhao et al. 2014; Sonta et al. 2017). A component selection process that utilizes a Variational Bayesian Gaussian Mixture Model (VB-GMM) determines the number of activity states present in the data. For the component selection process, a VB-GMM is applied for each occupant for each day separately, then a single number of components (M) is chosen for all occupants. After the component selection process, new Gaussian Mixture Models (GMM) are fit to the data for each occupant for each day separately, and the energy use data is classified based on the model fits.

The component selection process is based on a variational Bayesian inference method that utilizes the GMM as its basis. We employ this component selection process as a means of alleviating the need to make any *a priori* assumptions as to the number of activity states the occupant have. The component selection process makes use of engineering domain knowledge about occupant dynamics and plug load energy consumption in order to glean more compelling insights about occupant activities. In particular, we allow our method to recognize when data corresponding to occupant presence

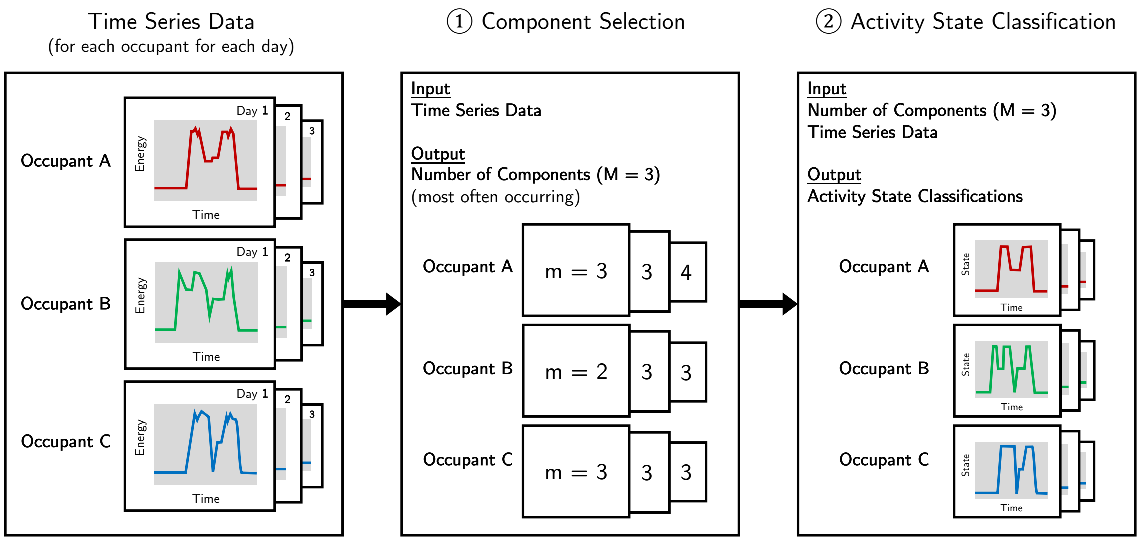

Figure 2.1: Conceptual outline of our classification method.

is classified as distinct from data corresponding to occupant absence, and we allow the component selection process to analyze the presence data separately. This flexibility, which would not be possible without incorporating knowledge about occupant dynamics, building energy use, and plug load energy data, allows our method more fully analyze the data. After the component selection process, new GMMs are fit for each occupant and for each day using the inferred number of components. This step ensures that all periods of occupant energy use are being classified with the same number of components, enabling effective comparison across occupants. This avoids situations in which, for example, one occupant tends to have 3 components, and another tends to have 4 components, and it enables extraction of insights across the entire occupant population in a building. Figure 2.1 shows a conceptual outline of our method, as applied to a sample dataset of three occupants over three days. For the purposes of this example, the number of components is arbitrarily chosen as 3.

2.3.1 Gaussian mixture model

We define a time series of plug load energy use measurements collected at the desk-level for each occupant:

$$\mathbf{X}_{i,d} = \{x_1, ..., x_T\} \tag{2.1}$$

where i is the occupant index (for all occupants $1, ..., I$), d is the day index (for all days $1, ..., D$), and T is the number of time steps in the period of study (T depends on the amount of time between measurements, which can be set as required by the user). For the purposes of this study, we utilize a time step of 15 minutes, resulting in $T = 96$ if the full day is analyzed. Data collected over one

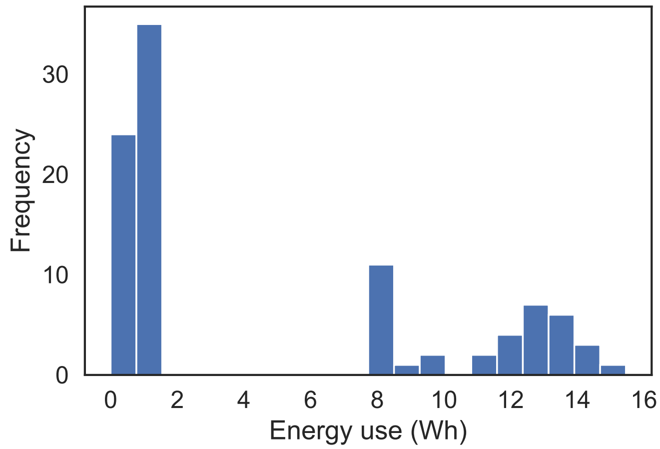

Figure 2.2: Typical plug load energy use histogram for one occupant over one day.

day for one occupant, as indexed by i and d (i.e., $\mathbf{X}_{i,d}$) defines a full sub-dataset. Figure 2.2 shows a histogram of a typical full sub-dataset. Many algorithms can be applied to the classification or clustering of granular (sub-hourly) time series datasets, such as k-means, naive Bayes, and support vector machines (Zhao et al. 2014). However, for these classifiers to be effective in determining occupant activity states, either they need to be trained on data that describes the true state of occupancy for each occupant in the study, or their parameters need to be tuned on existing data by the user of the method. Collecting this ground truth data for training is possible in a small experimental setting, but as previously stated, it becomes quickly intractable as we move to the analysis of large commercial buildings or even portfolios of buildings.

To derive a method to effectively classify energy states without training data, we looked to domain research on analyzing human activity states and dynamics. Previous work analyzing human heart-rate data has utilized a GMM to classify time series data into states describing some physical activity phenomena (Costa et al. 2012). Heart rate measurements taken over the course of a day exhibit multimodality due to the various physiological processes that affect heart rates. We have found plug load energy use data to exhibit similar multimodality, as can be seen in Figure 2.2, and thus we adopted the GMM as a basis of our method.

A GMM is based on the notion that the unimodal Gaussian distribution is useful and common in modeling real world data, and when observed data is clustered around multiple peaks—as is the case both with human heart rate variability and with occupant energy use—this multimodal data can be effectively modeled as a mixture of multiple unimodal Gaussian distributions that may or may not be independent. In a GMM, the likelihood function for observation x is given by:

$$p(x) = \sum_{k=1}^{K} \phi_k \mathcal{N}(x|\mu_k, \sigma_k) \tag{2.2}$$

where K is the number of mixture components and ϕ_k is the weight of each component. Each component follows a normal distribution with mean μ_k and standard deviation σ_k:

$$\mathcal{N}(x|\mu_k, \sigma_k) = \frac{1}{\sigma_k\sqrt{2}\pi} \exp\Big(- \frac{(x-\mu_k)^2}{2\sigma_k^2}\Big) \tag{2.3}$$

The values of each of the component weights (ϕ_k) sum to one so that the total probability distribution is normalized. Each component of the GMM corresponds to a *state* present in the data, with each state being described by the component's Gaussian distribution.

2.3.2 Variational Bayesian Gaussian mixture model

It is common to use Expectation Maximization (EM) to fit each component of a GMM, but this method requires the user of the model to make an *a priori* assumption about the number of components that are used in the GMM, and therefore if adapted to the occupant energy use setting, it would require the user to already know the number of energy use states naturally occurring in the office setting. However, requiring this prior greatly inhibits the extensibility of the model as it may be difficult to discern the number of energy use states prior to collecting data. As a result, we aimed for our model to be adaptable to the natural number of energy use states embedded in the data. We use a variation of the GMM called the Variational Bayesian GMM (VB-GMM) to allow the model to use Bayesian inference to choose the number of components. Recent developments in statistical inference have led to the accessibility of such models (Costa et al. 2012). We consider the GMM as described above, and for each observation x_t we include a corresponding latent variable $\mathbf{z}_i$ comprising a 1-of-K binary vector with elements z_{tk} for $k = 1, ..., K$. The observed dataset for each occupant can be denoted by $\mathbf{X} = x_1, ..., x_T$ and the latent variables by $\mathbf{Z} = \mathbf{z}_1, ..., \mathbf{z}_T$. Given the component weights $\boldsymbol{\phi}$, we can write the conditional distribution of $\mathbf{Z}$:

$$p(\mathbf{Z}|\boldsymbol{\phi}) = \prod_{t=1}^{T}\prod_{k=1}^{K} \phi_k^{z_{tk}} \tag{2.4}$$

We can also write the conditional distribution of the observed data, given the latent variables and component weights:

$$p(\mathbf{X}|\mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = \prod_{t=1}^{T}\prod_{k=1}^{K} \mathcal{N}(x_n|\mu_k, \Lambda_k^{-1})^{z_{tk}} \tag{2.5}$$

where $\boldsymbol{\mu}$ is the set of component means and $\boldsymbol{\Lambda}$ is the set of component precisions defined as the inverses of the standard deviations. When solving the model, we introduce priors over the parameters $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{\phi}$. Following common Bayesian statistical practices, we use a Dirichlet distribution over the mixing coefficients $\boldsymbol{\phi}$ and a Gaussian-Wishart prior governing the mean and precision of each component. We utilize the Python scripting language and the Scikit learn package (Pedregosa et al. 2011) to estimate the parameters of the Variational Bayesian GMM. The model flexibly chooses the number of final components that best describe the data out of a given possible number of components. In addition to the classification results, a key output of a VB-GMM fit is the number of components (n) that are used to classify at least one data point. A full theoretical description of the Variational Bayesian GMM can be found in (Bishop 2006).

2.3.3 Component selection

The component selection process utilizes the Variational Bayesian GMM applied to time-series plug load data. The overall component selection process is applied to all of the collected plug load data across the space being analyzed. The process utilizes the VB-GMM model by applying it to each *full sub-dataset* independently. Figure 2.3 shows the flow of the component selection process. First, the set of data points in the period of analysis is defined. Then, a Variational Bayesian GMM is fit to each sub-dataset—that is, to each occupant and each day independently—inferring the number of components (n) present in the sub-dataset for one occupant for one day. As long as there are more days to analyze for an occupant and more occupants in the space being analyzed (more sub-datasets), this process is repeated, until all occupants and days are analyzed. After all occupants and days are analyzed, the number of components most often chosen (the mode of the full set) is determined as M.

In this first step, all of the data points in each day are analyzed for component selection, and we refer to this process as the primary component selection process. Additionally, we refer to the number of components chosen by the method in the primary component selection process as M_1. We define three possibilities:

1. $M_1 = 1$. If only one state occurs most often, only one state is present in the data and no further analysis is done.

2. $M_1 = 2$. Often, the VB-GMM selects two states most often in this primary component selection process. Figure 2.4 shows a typical occurrence of the VB-GMM choosing two components for a full sub-dataset. By observing the data, we note that this classification essentially splits plug load energy use data for each day and occupant into two states: (1) a state in which very little or no energy is being used at the workstation, and (2) a state in which at least some energy is being used at the workstation. As can be inferred from Figure 2.4, the higher energy state as classified by the GMM has much more variability than the lower energy state. This typically occurs because it is separating data associated with presence from data associated with absence. Previous analyses of occupant activities have adopted the practice of first separating the data corresponding with occupancy from that corresponding with absence and analyzing the occupancy data separately (Milenkovic & Amft 2013). Therefore, we integrate this domain knowledge on occupant dynamics into our method: when exactly two components

are chosen in the primary component selection process, the data classified in the higher energy component for each occupant is separated and a new VB-GMM is applied to this separated data (as long as two or more data points are in the higher component). We define this process as the secondary component selection process, and the process in Figure 2.3 is repeated, this time just for the data classified in the higher energy component. Again, we determine the number of components (n) chosen most often across all occupants and all days, and refer to this value as M_2. After the component selection process, in the subsequent classification step, this two-step process is once again applied.

3. $M_1 > 2$. If three or more components occur most often in the primary VB-GMM, this number of components is used in the classification step.

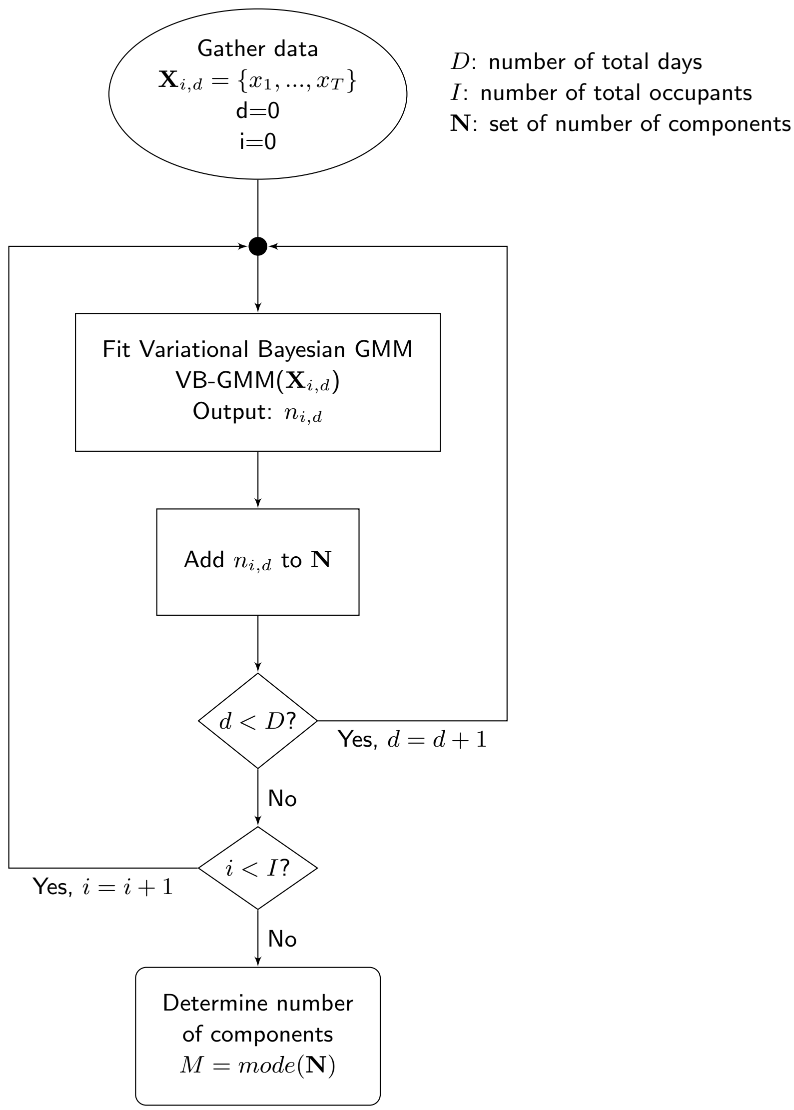

Figure 2.3: Component selection process using the Variational Bayesian GMM.

2.3.4 Activity state classification

Once the number of components is inferred through the component selection process, we fit new GMMs to the data so that all occupants are classified into one of the same number of activity states. GMMs are fit independently for each occupant for each day, but all GMMs use the same number of components, as determined in the component selection process. Once a GMM is fit to data, the data points X are classified into components: $\{x_1, ..., x_M\} \in X$. Figure 2.5 shows the flow for the activity state classification process applied to data from one occupant for one day. If the primary

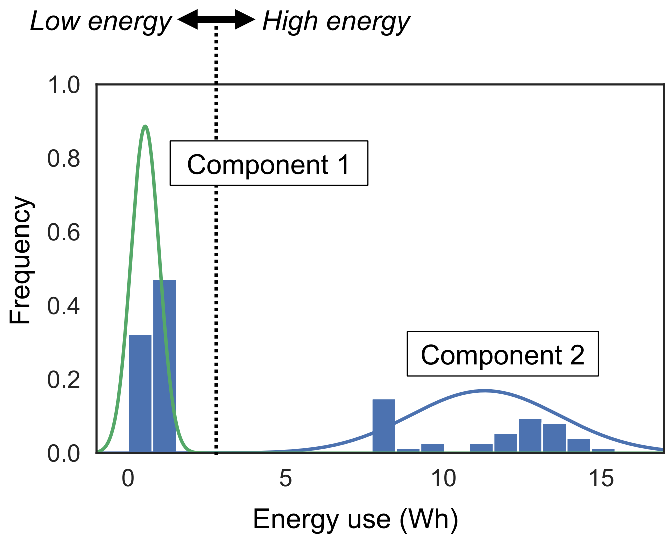

Figure 2.4: Result of fitting 2-component GMM to all energy use data for one occupant over one day, with dashed line showing where the "Low energy" and the "High energy" components cross.

component selection process resulted in exactly two components ($M_1 = 2$)—one corresponding to occupant absence (x_{low}) and the other corresponding to occupant presence (x_{high})—then a GMM is fit with two components to the data and the higher energy component is separated. Then a new GMM is fit to the higher energy data based on the number of components selected in the secondary component selection process (M_2). If the primary component selection process resulted in three or more components ($M_1 > 2$), a GMM is fit to the original data with that number of components. Once the GMMs have been fit to the data for each occupant for each day separately, the data points are classified into their respective activity states as determined by the GMM fits.

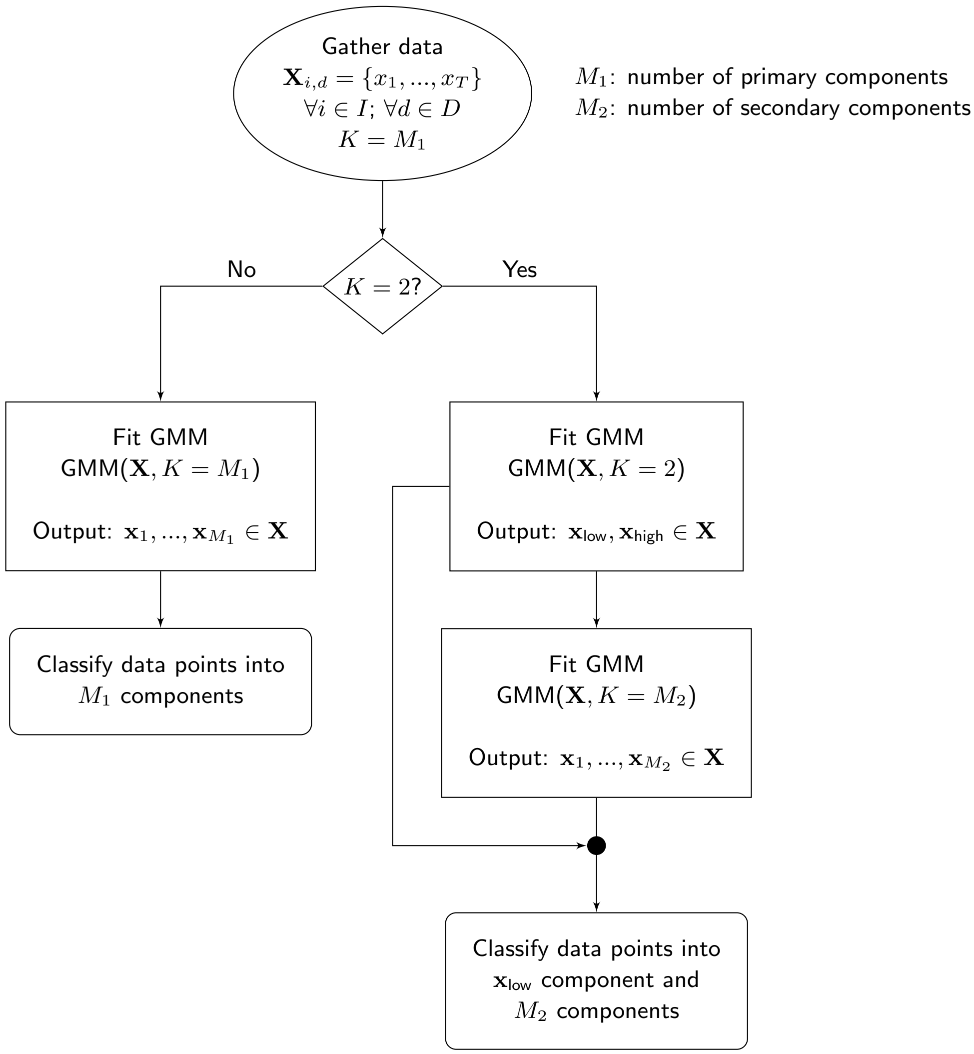

Figure 2.5: Activity state classification process using the GMM for one occupant and one day.

Figure 2.6 summarizes the classification process for one occupant and one day in the case where the primary component selection process resulted in two components, and the secondary component selection process also resulted in two components. In this case, a total of three activity states are utilized to describe the data. Figure 2.6a shows a fit of a 2-component GMM to an occupant's energy use data over one day. Any energy use value can be assigned probabilities that it is associated with each component of the GMM. The point between the means of the two components where the probabilities are equal—the point in Figure 2.6a at which the probability density function lines cross—becomes the cutoff point between the lower energy (x_{low}) and higher energy (x_{high}) states: any value below this point is part of the lower energy state and any value above this point is part of the higher energy state. Each of the occupant's energy use values observed throughout the day is

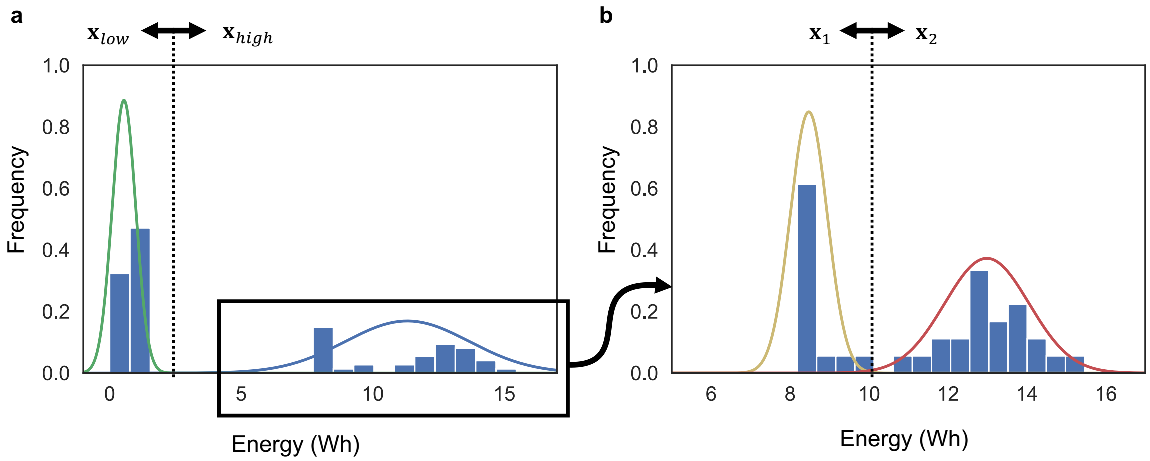

Figure 2.6: Activity classification process when primary component selection process results in two components. (a) GMM fit to all energy use data from one day. (b) A second GMM fit to just the "higher energy" data from (a).

classified according to this system. In the secondary step, the higher energy state data is separated out, and another 2-component GMM is fit to just this data, as shown in Figure 2.6b. Once again here, the point at which the probability density function lines cross becomes the cutoff point between the two states. As a result of this specific process, three states are determined (x_{low}, x_1, x_2), which can be visualized by the green, yellow, and red density functions in Figure 2.6.

We design our method to automatically determine when the *secondary component selection process* is necessary in order to provide deeper insights into the data associated with occupant presence. We do this by incorporating explicit domain knowledge into the method. Specifically, the method recognizes when the GMM is only forming two clusters—one associated with absence, and one associated with presence—and it then analyzes the more variable presence data in a second step. If we did not design our method with this functionality to recognize when two steps are necessary, the analysis of the data would be much less insightful, limiting the final number of activity states to two—presence and absence—in some situations. Without incorporating domain knowledge about the typical structure of plug load data signatures, our method would be less effective in describing behavioral patterns of occupants and in turn be limited in its applicability for informing building design and operations.

2.4 Validation study

In order to validate and benchmark the performance of our proposed energy state classification method, we conducted a two-week study involving seven occupants in an academic office setting. The occupants were five graduate students, a postdoctoral research scholar, and a professor in a total of three offices. Each occupant had his or her own desk. Three graduate students and the postdoctoral research scholar shared one office. Another two graduate students shared another office, and the professor had their own office. We note that validation studies of this size are common in the field of occupant activity recognition in office settings (Ahmadi-Karvigh et al. 2018). Typical occupant activities included working on a computer at a desk, going to classes, holding meetings in one of the three offices, and taking lunch, coffee, or social breaks.

2.4.1 Data collection & analysis

Plug load energy use over 15-minute intervals was monitored and stored using HOBO data loggers connected to a power strip at each occupant's workstation. Each station consisted of a computer plugged into the power strip, and five of the seven workstations had a monitor plugged into the power strips. Participants also were allowed to include some miscellaneous but relatively small power loads, such as cell phone chargers.

To determine how well our classification model captures activity states of occupants, the participants of the validation study manually recorded their activities and the times of their activities over the course of each day. Occupants kept track of the times they arrived and left their desks both at the start/end of their workday and during breaks and meetings throughout the day. This recording of occupant activities formed the "ground truth" data of occupant activity states, with noted activities indicating shifts between the states of occupancy defined in our methodology. To compare the results of our activity classification method with the ground truth data, we counted each instance of an occupant arriving at or leaving his or her desk as an "event" associated with a transition between states. For example, if the occupant notes an arrival at his or her desk at the start of the workday, this event is associated with a shift from a lower energy state to a higher energy state. Similarly, if the occupant notes that he or she has left the desk, this event is associated with a shift from a higher energy state to a lower energy state. Over the course of the two-week study, 345 events were recorded by the seven occupants.

We input the collected plug load energy use values to our classification method and compared each day's state classification results with the ground truth activity data. (The classification process resulted in a total of three states.) If the occupant's recorded event corresponded with a shift between states that aligns with the activity (e.g., a shift from a higher state to a lower state if the occupant leaves his or her workstation), then we consider this event to be correctly classified (i.e., a true positive). If the method shows a shift between energy states, but no event was recorded, we consider this a false positive. And if the occupant records an event but no shift between states was detected by the method, we consider this a false negative. Figure 2.7 shows the raw data and the occupant activity detection results for one occupant over a workday, with the two most common of these scenarios—true positives and false negatives—depicted by annotation. At 9:07 a.m., the

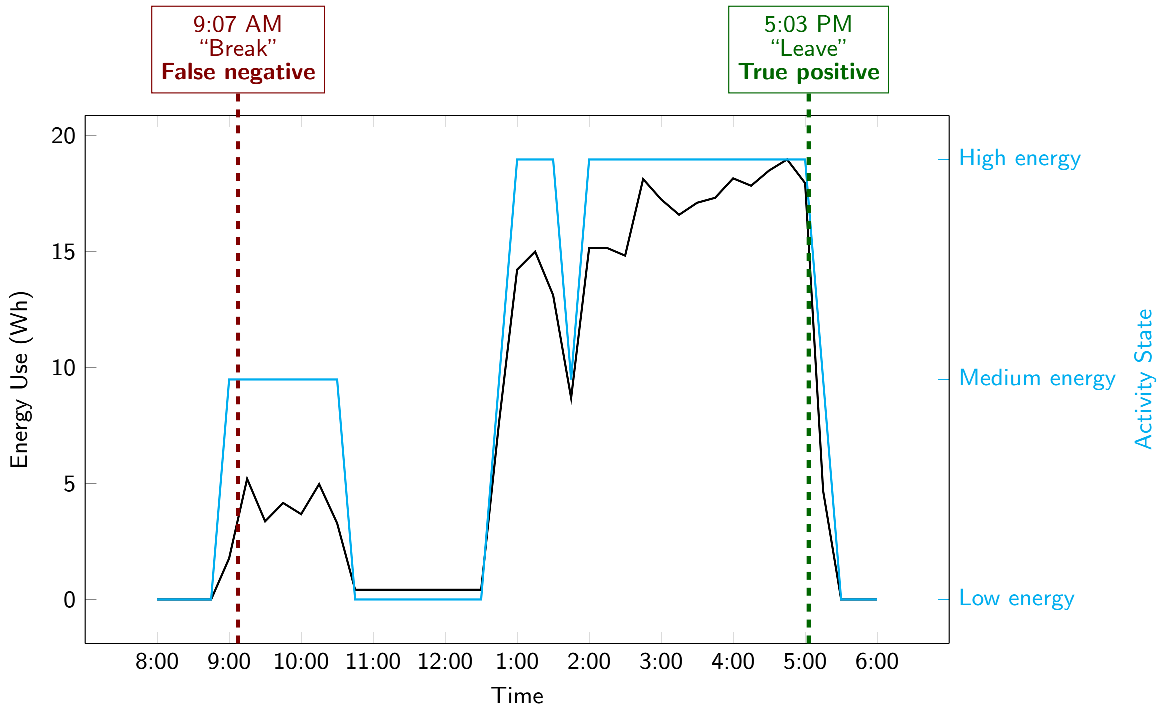

Figure 2.7: Raw energy use values and classification results, shown with an example of a false negative as well as a true positive.

occupant noted a break, but the method did not indicate a shift in occupancy states at this time—hence, we annotate this as a false negative. At 5:03 p.m., the occupant notes that he or she left for the day, and the method indicated a corresponding shift from a higher state to a lower state—hence, we annotate this as a true positive.

2.4.2 Validation results & discussion

We repeated the analysis shown in Figure 2.7 for each occupant for each day to understand the effectiveness of our classification method. For building management systems to be able to effectively implement information about the behavior of occupants, a high level of accuracy is required. However, not all measures of accuracy have the same implications. As indicated in [4], situations in which detection sensors or methods determine a state of absence in the building when in fact the building is occupied are more problematic than when they determine a state of occupancy when in fact the building is unoccupied. These situations can lead to lights switching off or a lack of HVAC service to occupants in a room, which can not only cause discomfort, but it can also lead to occupants overriding the intelligent building systems that make use of the information about occupant presence (O'Brien & Gunay 2014). Consequently, it is extremely important to minimize the rate at which these sensors determine false absence. At the same time, however, if the rate at which sensors determine false occupancy is too high, building managers will not be able to effectively make use of the information, leading to missed opportunities for energy efficiency. Recent work (Gunay et al. 2016; Erickson et al. 2009) has suggested that in order to be effective this false occupancy rate can be only as high as 20%.

To measure the performance of our method in capturing occupant activity shifts, we utilize the precision and recall metrics. The precision metric for the method—the ratio of true positives to all positives (including true and false)—is calculated to be 92.7%. The recall metric—the ratio of true positives to the sum of true positives and false negatives—is calculated to be 73.9%. Both of these metrics suggest that our method is accurate in capturing changes in occupant activity states.

It is important to note the rates at which our method leads to false measures of absence versus false measures of occupancy. We can infer from Figure 2.7 that our definition of false negatives typically occurs when an occupant indicates a change in state associated with a break from working at their workstation, but the method does not recognize this change in state. (It is also possible that a false negative could be associated with a missed state transition associated with starting the workday or leaving the workday, but these situations occur very rarely.) Therefore, a false negative is typically associated with a false measure of occupancy at the workstation (i.e., the occupant has taken a break but the method has not recognized this break). Alternatively, a false positive indicates that the method has detected a change in state when no state change has been recorded by the occupant. We note that this is not the same type of error as the false negative that is indicated in Figure 2.7, where the occupant did record an activity, but the algorithm did not detect a change in state. Most often, false positives are associated with the method detecting that the occupant has taken a break, when in fact the occupant is still working at the workstation. Therefore, it is very important to minimize these false positives, because false measures of absence are more problematic than false measures of occupancy, as discussed above. Our results show a very high precision rate of 92.7%, indicating that false measures of the absence of occupants are minimized. At the same time, our recall rate of 73.9% suggests that our method is still providing valuable information about the true state of occupant activities in the building. We also determined the overall false occupancy rate by comparing the output of the method with the occupant notes. We find the overall false occupancy rate to be 5.8%—well below the 20% threshold suggested by the literature (Gunay et al. 2016; Erickson et al. 2009).

Table 2.1 shows the precision and recall metrics for each occupant in the validation study. As shown in the table, Occupant 7 has the lowest values for both precision (84%) and recall (44%). We hypothesize the precision and recall discrepancies occur because Occupant 7 has atypically low variability in actual energy use. The standard deviation of energy use for Occupant 7 for medium and high energy states was 2.82 Wh, whereas the standard deviation of energy use for the other occupants ranged between 4.41 and 9.74 Wh. Therefore, in raw energy consumption terms, the energy consumed by Occupant 7 was less likely to change significantly in correlation with a noted event. On days where the method performed abnormally poorly (i.e., more than 50% of noted events were false positives), the average standard deviation of energy use values was 3.56 Wh. On days where the method performed relatively well (i.e., less than 25% of events were false positives), the average standard deviation of energy use values was 6.79 Wh. This indicates that higher variability in energy use values is correlated with better performance of the model.

Table 2.1: Precision and Recall metrics for each occupant in the validation study.

Occupant	Precision	Recall	False Occupancy Rate
1	100%	77%	5.8%
2	93%	85%	2.6%
3	93%	80%	1.6%
4	91%	82%	4.2%
5	99%	60%	17.1%
6	97%	81%	2.1%
7	84%	44%	19.8%
All Occupants	100%	73.9%	5.8%

2.5 Case study: San Francisco office building

We applied our method to data collected in a typical office building in San Francisco, CA, USA in order to demonstrate the merits of our methodology for informing energy efficient operations and space utilization of a commercial building. The occupants were employees of an organization headquartered in San Francisco. Based on observations during a site-visit, we found that activities in the space were comprised of typical office work at a computer workstation, meetings, and breaks from working (including lunch and social breaks).

2.5.1 Data collection

We utilized plug load data collected for 47 occupants in an open-office building in San Francisco, CA, USA using Enmetric plug load sensors (*Enmetric Systems* n.d.). The sensors continuously collected and reported total energy use over 15-minute periods at the individual desk level. By manual inspection of the data, we found that some of the sensors had connectivity issues, and that on certain days, these sensors did not report any energy consumption at all. We therefore limited our analysis of this data to a clean segment of the data spanning 41 continuous days. We note that our method could be applied to data outside this time span, but that more occupants would simply be classified into a very low energy state throughout the day. For the sake of brevity and clarity, we choose to concentrate our analysis and discussion the 41-day period that had consistent data.

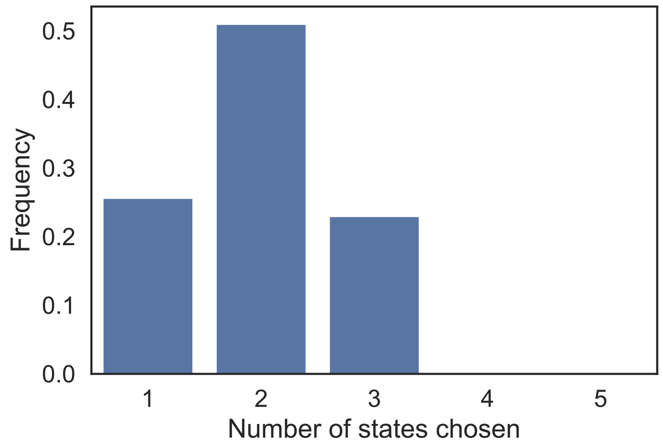

Figure 2.8: Number of states chosen (out of 5) by Variational Bayesian GMM for all energy use values for each occupant for each day.

2.5.2 Case study results & discussion

Following the methodology described in section 2.3 above, we first performed the component selection step and then used the results to classify the data into activity states. During the component selection process, each Variational Bayesian GMM was limited to using up to 5 components to classify the sub-dataset. With a total of 1,927 sub-datasets, the Variational Bayesian GMM was applied to each sub-dataset, and the majority of the models (51.3%) chose 2 components to classify the data, as shown in Figure 2.8. None of the models chose more than three components to classify the data. After determining that the Variational Bayesian GMM most often chose 2 components for this dataset, we fit another GMM to the plug load energy use data for each occupant and for each day, using exactly 2 components for each GMM. This process allows for the classification of each 15-minute period of occupant activity into one of 2 states: *low energy* and *high energy*. Because exactly two components were chosen in the primary component selection process, the data classified into the *high energy* state was separated and the secondary component selection process was performed.

Another Variational Bayesian GMM was fit to each day's high energy data for each occupant, and once again the majority of the models (58.7%) chose 2 components, as shown in Figure 2.9. As long as the initial *low*/*high* energy state classification results in at least two data points classified as *high energy*, we can then fit another 2-component GMM to the data points initially classified as *high energy*, reclassifying the *high energy* state into two states: *medium energy* and *high energy*. If one or fewer data points is initially classified as *high energy*, our method does not perform the second classification into *high* and *medium* energy states.

Our method resulted in a total of three possible states for each 15-minute period for each day for each occupant, which we delineate as *low energy*, *medium energy*, and *high energy* states. By

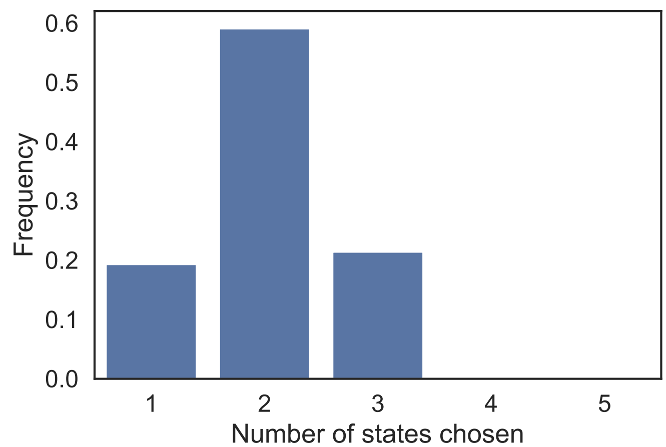

Figure 2.9: Number of states chosen (out of 5) by Variational Bayesian GMM for just the "high energy" energy use values for each occupant for each day.

allowing the Variational Bayesian GMM to adaptively fit to the plug load energy use data for each day and occupant, we infer the natural number of states that describe the overall dataset. The result of this analysis is a mapping from raw energy use values to states of occupant activities in the building. Figure 2.10 illustrates this mapping for each occupant over the floor plan of the building. As Figure 2.10a shows, it can be difficult to understand the meaning of the raw energy use values, since a value of, for example, 5Wh over the 15-minute period can mean very different things for two different occupants. However, as Figure 2.10b shows, the mapping into activity states abstracts information about these energy use values, providing deeper insights into the activities of occupants across the floorplan of the building.

A potentially useful application of this methodology is to aid in understanding nuanced aspects of the energy efficient operations and space utilization of buildings to which it is applied. Figure 2.11 shows the progression of occupancy states for each of the 47 occupants over one afternoon/evening in the study period, from the 15-minute period ending at 5:00pm to the 15-minute period ending at 7:15pm (10 time-steps in total). Monitoring this progression provides insight into how the space utilization in this office setting changes over time. This information provides value for building managers who seek to ensure that spaces are being utilized efficiently. Furthermore, as building designers and engineers embrace the notion of *performance based design*, they are beginning to adapt the programming phase of the design process to include more nuanced understandings of space utilization. Monitoring and visualizing occupant activity states in buildings can help in the development of space use frameworks that more closely align with the true states of occupant activities in buildings.

As building lighting and HVAC systems continue to be installed with more granular control over spatial and temporal dimensions, building managers can also use this information to optimize the control of these building systems to reduce the amount of energy required to provide services to their

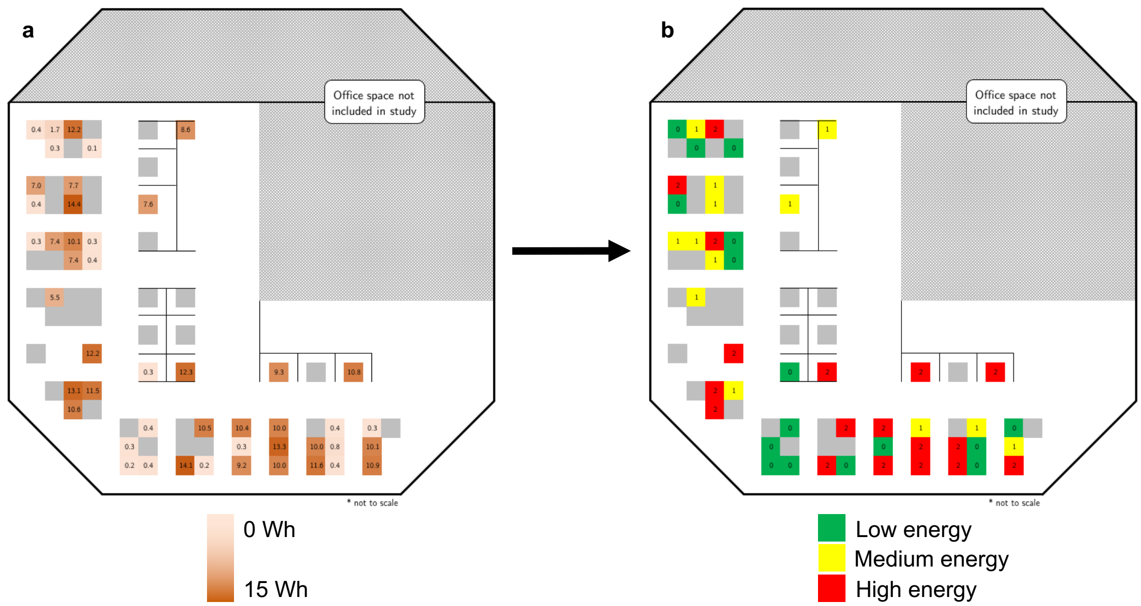

Figure 2.10: Mapping from raw energy use values (a) to occupant activity states (b) for each occupant over one 15-minute period

occupants. Visualizing this information also provides for the opportunity to make recommendations for co-optimization between occupants, space, and building systems. For example, if groups of occupants who do not have desks near each other regularly shift to *low energy* states at the same time, these occupants could potentially be relocated to be physically near each other so that building systems can reduce services like lighting and HVAC in the space they occupy. An example of such a realignment strategy is depicted in Figure 2.12. Here, occupants within the blue circles are identified as occupants that shift from a higher state to a lower state from 2:00pm to 2:30pm on a specific day in the study period, perhaps for a meeting or to take a break at the same time (Figure 2.12a). If this pattern recurs commonly in building, one potential strategy would be for these occupants to move to workstations that are physically near one another (Figure 2.12b). After realignment, lighting and HVAC systems could adjust to the change in occupancy states at the identified workstations.

The analyses presented in Figure 2.11 and Figure 2.12 demonstrate the ability of our model to provide new knowledge to building managers and designers. Without requiring training data or an *a priori* assumption about the number of activity states in the plug load data, we are able to glean effective insights about occupant activities in the space. Furthermore, the visual representation of the occupancy states across the floorplan offers new insights that could not otherwise be interpreted from the raw plug load energy use values. As Akbas et al. (2007) notes, effective visual representation of new spatio-temporal information can be an effective decision-support tool for managers and designers. As a result, our method has the ability to help building operators make decisions for

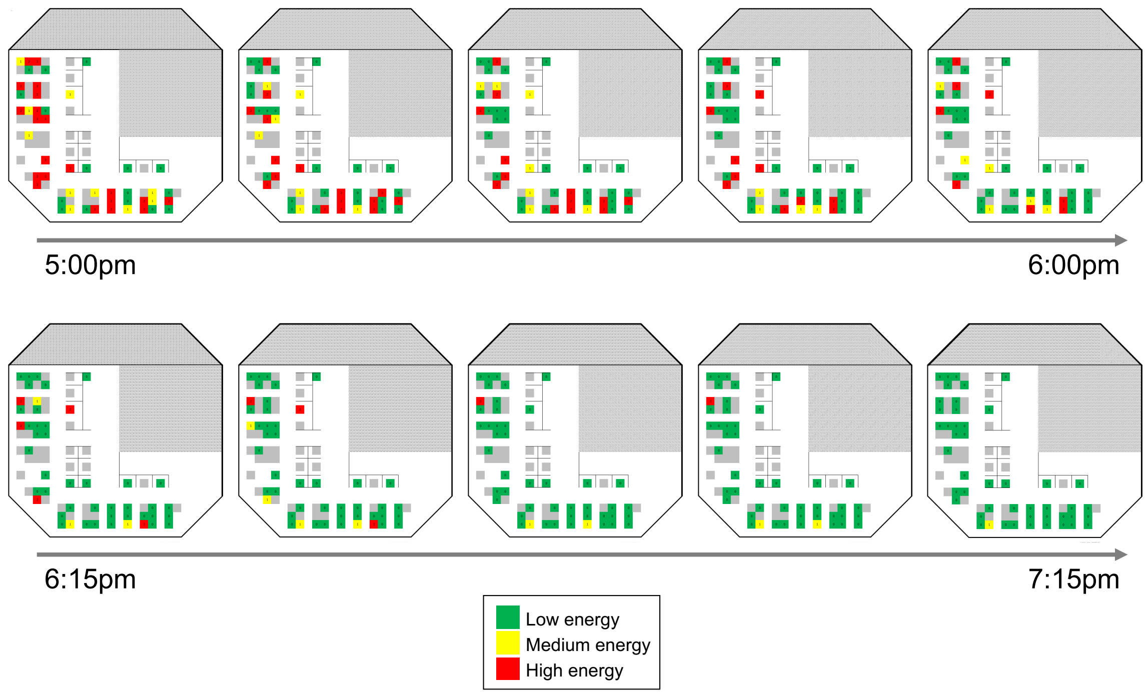

Figure 2.11: Visualization of the changing space-use levels over time using activity state classification method.

energy efficient system management and to help designers build models of occupant activities for improved design of future spaces.

2.6 Limitations and future work

The main limitations of our method stem from the inherent constraints of plug load energy use data to capture activities of occupants in buildings. While the method performs well enough to provide valuable information to building engineers, operators, and designers—based on suggested precision metrics from the literature—there is opportunity to further improve the precision. For example, while plug load energy use data typically changes when occupants take extended breaks from their workstations, there are situations in which plug loads stay high while occupants take short breaks from their workstations. Future work could incorporate the use of other sources of data, such as infrared sensors, in order to complement the plug load data collected for our method. A composite data stream that includes multiple sources could lead to more precise detection of occupant activities.

In addition to possible improvements in accuracy, future work could consider identifying occupant activities that are not associated with plug load energy use. Plug load sensors are cost-effective for this task and often easily accessible, since they are commonly installed in office buildings for

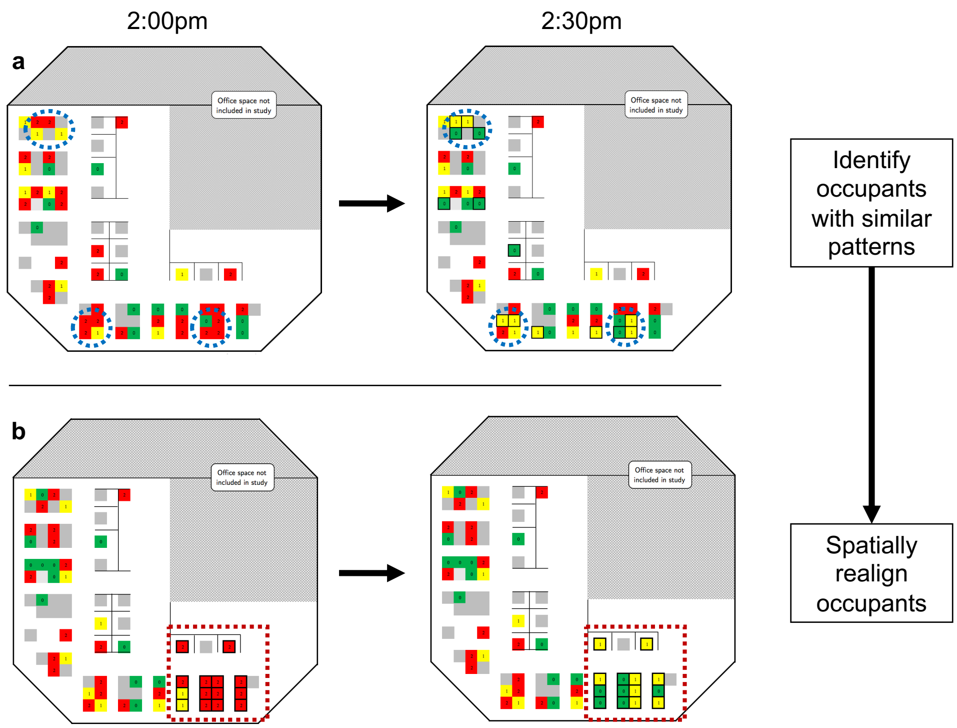

Figure 2.12: Potential occupant realignment strategy. (a) Occupants with similar patterns are identified. (b) After realignment, intelligent building systems can take advantage of activity state shifts.

various purposes beyond inferring activity patterns, such as for monitoring the energy consumption of miscellaneous equipment. While their data provide a good proxy for occupant activities, a more robust understanding of occupant behavior should include non-energy-intensive activities. Again, such activities could be recognized using data streams that are complementary to the plug load energy use data. Similarly, the method developed in this paper could be extended for the analysis of analogous data produced from other sensors, in particular when time series data exhibits multi-modality and domain knowledge about the states associated with the components of the distribution is known.

It is important to note that while the validation study discussed in Section 4 shows reasonable reliability and internal validity, claims about its external validity must be made with caution. While our inference method is designed to be able to adapt to individualized settings, where different occupants have different baseline, average, and maximum plug load energy consumption, further studies are needed to determine the validity of the classification results in settings beyond this internal validation study. Our internal validation study demonstrates the robustness and adaptability of our method within the setting of the study, but future studies with large-scale ground truth data collection are needed for broader claims about the true scalability of our method.

One exciting area of future work involves utilizing this data to gain insight into the natural structure of the occupant network in the building. Our method provides information about states of occupant activities in the building, which can be useful in understanding not just individual activity states, but also the relationships among occupants. For example, two occupants that have very similar patterns of activities in buildings could be highly related socially or organizationally. Building managers and designers could make use of this information by potentially suggesting shifts in the occupant layout in the building, allowing building systems to be more closely aligned with the states of occupancy across the building. Furthermore, by gaining an understanding of the structure of the network of occupants, eco-feedback systems that attempt to convince occupants to adjust their behavior could become more effective, as the network structure of occupants has been shown to have high importance in these strategies (Anderson et al. 2014; Jain et al. 2012).

2.7 Conclusions

The main contributions of this work are twofold: first, to introduce a new adaptable method that integrates knowledge-based and data-driven approaches for inferring occupant dynamics in building; and second, to demonstrate how our proposed method can be utilized to infer the occupant dynamics occurring in a building and inform intelligent optimization strategies for energy efficiency and space utilization. By integrating a variational Bayesian version of the Gaussian mixture model with explicit domain knowledge about occupant dynamics and plug load data signatures, we designed our method to require no ground truth data to perform with a high level of accuracy. These methodological design decisions allow our method to be more easily applied to situations where ground truth data is difficult to collect, such as when there are many occupants across one or more buildings.

In analyzing newly accessible plug load data streams to infer occupant activity states in buildings, we can gain a deeper understanding of the complexity of occupant dynamics at a high level of spatial and temporal granularity. In turn, this deeper understanding translates to new knowledge about occupant dynamics that can help building designers, engineers, and managers better understand how occupants respond to the spaces they occupy. These decision-makers will now be armed with the knowledge that can enable them to intelligently manage building operations and design to enhance energy efficiency, space utilization, and occupant satisfaction. Buildings will inevitably continue to play a crucial role in each of our lives as occupants and in the world's sustainable energy future. New methods that combine the extant knowledge of occupant dynamics and building systems with emerging data-driven methods could provide us with the necessary insights to design, operate, and manage the next generation of high-performance buildings.

Chapter 3

Learning socio-organizational network structure in buildings with ambient sensing data

This chapter is adapted from the following paper: Sonta, A. and Jain, R. K. (2020). "Learning socio-organizational network structure in buildings with ambient sensing data," *Data-Centric Engineering*, 1, E9.

Abstract

We develop a model that successfully learns social and organizational human network structure using ambient sensing data from distributed plug load energy sensors in commercial buildings. A key goal for the design and operation of commercial buildings is to support the success of organizations within them. In modern workspaces, a particularly important goal is collaboration, which relies on physical interactions among individuals. Learning the true socio-organizational relational ties among workers can therefore help managers of buildings and organizations make decisions that improve collaboration. In this paper, we introduce the Interaction Model, a method for inferring human network structure that leverages data from distributed plug load energy sensors. In a case study, we benchmark our method against network data obtained through a survey and compare its performance to other data-driven tools. We find that unlike previous methods, our method infers a network that is correlated with the survey network to a statistically significant degree (graph correlation of 0.46, significant at the 0.01 confidence level). We additionally find that our method requires only 10 weeks of sensing data, enabling dynamic network measurement. Learning human network structure through data-driven means can enable the design and operation of spaces that encourage, rather than inhibit, the success of organizations.

3.1 Introduction

The structure of an organization can be described by a network of relational ties among its individual members. This network is often portrayed as a simple chart representing hierarchy, but in reality, it is a complex web of interactions that enables the sharing of information, work-related advice, and personal advice among members of the organization and across levels of hierarchy (Krackhardt & Hanson 1993). These socio-organizational relationships manifest themselves both digitally and physically, though research continues to show that physical interactions are the cornerstone of collaboration and meaningful ties (Waber et al. 2014). Because of the importance of these physical exchanges, the design of the physical workspace—the commercial building—can have an impact on the nature and frequency of interactions among members of organizations (Sailer & McCulloh 2012), which can be critical to the success of the organization (Kabo et al. 2015). However, learning the true social and organizational ties among workers, which would be helpful in designing and managing these spaces, remains a challenge. Fortunately, recent advances in physical sensing strategies provide a pathway for inferring such relational ties by analyzing interactions that happen in physical space. The core investigation in this paper, therefore, is the inference of socio-organizational relationships of office workers using ambient sensing data—in this case, plug load energy sensors installed at the desk level—that capture patterns of human use of space within the building.

The problem of learning organizational network structure has been well studied due to its importance to the fields of management science and organizational theory (Krackhardt & Hanson 1993; Tichy & Fombrun 1979). More traditional, well-established methods born out of social science that make use of surveys, interviews, and observations have been widely used for this problem (Krackhardt & Hanson 1993; Marin & Wellman 2016; Tichy & Fombrun 1979); however, we identify two shortcomings from these approaches. First, simply measuring the structure of an organization through its internal structure of leadership and teams can miss other important characteristics of an office network, such as social relationships and relationships formed through spatial configuration, whereby increased proximity and shared use of space can drive interactions (as discussed in (Sailer & McCulloh 2012)). Second, survey approaches require considerable time and effort to administer. This cost can become intractable when measuring extremely large networks, such as those for large organizations that inhabit sizeable buildings or campuses. In order to capture the subtleties of workplace ties more fully, data-driven methods are increasingly being used to conduct complex analysis of workplace behavior. Data-driven methods can offer accurate and subtle insights into the nature of occupants' activities and utilization of spaces over time, thereby enabling the automated inference of socio-organizational relationship structure of office workers.

Recent developments in time series analysis have enabled inference of correlations and causal relationships among entities, including network structure of complex systems (Runge et al. 2019). However, the statistical and data mining tools that have been proposed for network inference are typically designed for non-social systems, such as biological, physical, or abstract networks (Friedman et al. 2008). Some recent work has adapted these more general methods for inferring social networks from time series data on human activity (Pan et al. 2012). Separately, recent research has shown that game-theoretic approaches can successfully construct social networks based on features ascribed to individuals (Yuan et al. 2018). In reviewing the literature, we identified two models that may be appropriate for the problem of learning socio-organizational networks from sensor data: the Graphical Lasso (Friedman et al. 2008) and the Influence Model (Pan et al. 2012). The Graphical Lasso, a model for estimating the inverse covariance matrix among entities producing time series data, was originally introduced in the context of learning protein interactions in biological systems (Friedman et al. 2008). However, more recent extensions have argued that the model may be applicable to social systems (Hallac et al. 2017). Conversely, the Influence Model leverages a generally coupled Hidden Markov Model, which produces an "influence matrix" that can be used to define a network. The Influence Model was originally introduced in the context of the electrical grid (Asavathiratham 2001), but has since been adapted to other problem settings, including the context of modeling influence among people in social settings (Pan et al. 2012). Because the literature has suggested that these data-driven models may be appropriate for modeling social networks, we adapted them to our unique problem setting involving ambient sensing data of building utilization. However, the use of ambient sensing in physical office spaces remains a key gap in social network inference. By leveraging this new paradigm of sensor deployment and data availability, we will be able to understand the operation of organizations and the spaces they inhabit through this lens of human activity in real time.

Our work here adopts network inference methods to infer the human network structure of office workers from distributed plug load energy sensors. These sensors are becoming ubiquitous and, as discussed in our previous work, can be used to model individual activity states at the desk level (Sonta et al. 2018). As we discuss in detail below, analysis of these activity states reveals times when office workers have the opportunity to interact with one another, a key component that drives collaboration and innovation for organizations and companies.

3.1.1 Organizational success, spatial layout, and network structure

The success of an organization can be described in part by the quality of the work performed within it. This notion of improved quality of work is difficult to measure and necessarily different for different workplaces. A company that prides itself on creativity and innovation is likely to care more about, for example, new ideas generated per day than number of words typed per minute. Despite the nebulous nature of organizational performance, research has pointed to key metrics that are particularly important in many of today's workplaces. Particularly in what researchers have dubbed the "knowledge industry"—organizations that trade in technology and human capital (Powell & Snellman 2004)—one key component of organizational success has crystallized as vitally important: collaboration (Soriano & Huarng 2013). The fields of economics and organizational theory have argued that complementarities among individuals are key to organizational success, which enables the view that collaboration and interaction among members are strong components of overall productivity (Ethiraj & Garg 2012). When considering the measure of collaboration among employees, understanding the structure of their relationships can give a manager a sharper sense of opportunities for collaboration as well as the ability to design interventions to improve it (Olguin Olguin & Pentland 2010).

The physical spaces of most organizations are office buildings. These buildings are created to meet the needs of their occupants, including needs relating to the subjective experience of the building as well as the indoor environmental quality (D'Oca et al. 2017). Architectural researchers have claimed that the organization of space is driven by the ordering of relationships among people (Hillier & Hanson 1984). This view creates the opportunity to leverage knowledge of the organizational structure in spatial design for collaboration. Moreover, researchers focusing on energy efficiency in buildings have noted that the structure of the occupants' social network has an impact on the way in which energy-efficient behaviors take root and spread in a building (Anderson et al. 2014). Connecting research on office building design and operation with research on organizational performance demands the need for a way to infer the complex relationships of office workers—the central question we investigate in this paper.

In the domain of workspace and organizational theory, researchers have noted a strong relationship between office design/layout and occupant satisfaction and performance (Kabo et al. 2015; Sailer 2011; Sailer & McCulloh 2012). Recently, researchers have noted that the physical design of buildings can have large impacts on different metrics related to productivity, such as communication, collaboration, creativity, and innovation (Kabo et al. 2015; Sailer 2011). Using the language of space syntax (Bafna 2003), researchers have defined metrics defined by physical layout and correlated them with occupant outcomes. For example, Peponis et al. (2007) found that higher levels of a single workspace's spatial integration correlated with more central positions in the organizational network for the individual occupying that workstation. Kabo et al. (2015) found that higher path overlap among occupants correlated with more successful collaborations. Generally, research has found that higher spatial relationships (e.g., proximity) improves the way individuals communicate and collaborate with one another in a building (Claudel et al. 2017; Kabo 2018; Wineman et al. 2009; Housman & Minor 2016). An accurate understanding of true relationships among occupants can be a critical tool in understanding the nature of work in buildings, and ultimately for suggesting spatial shifts that improve office worker performance and collaboration (Olguin Olguin & Pentland 2010; Sailer et al. 2015).

The network of relational ties among individuals in an organization also has implications for the physical performance of the building they occupy. Recently, researchers have found that understanding the network structure of building occupants can be useful in suggesting building layout changes that could reduce energy consumption of heating, cooling, and lighting systems (Sonta et al. 2017), as well as impact the efficacy of information campaigns aimed at targeting energy-efficient behavior, as discussed above (Anderson et al. 2014). While building energy performance has important environmental and societal implications, managers of organizations and buildings are typically driven to maximize the performance of the workforce. For the University of California system—whose operating budget is public data—total employee salaries, wages, and benefits constitute the vast majority of total expenditure. Compared to utility costs, employee costs were roughly 72 times more expensive in 2017–18, underscoring the notion that organizations are rational if they prioritize the productivity of their workforce over building energy efficiency (*Revenue and expense data, University of California* n.d.). With this economics driving decision-making, managers may be unlikely to make changes to the design of a building if they worry that adopting changes could disrupt productivity. Understanding the structure of the socio-organizational occupant network offers an opportunity to suggest design perturbations that can save energy without disrupting the natural flow of information, advice, and other components of work. Gaining insight into this network can enable new methods for co-optimizing these tightly coupled human and building systems that are fundamentally intertwined.

In this paper, we develop a method for automatically inferring the human network structure of office workers using ambient sensing data. Our method defines "opportunities for social interaction" as times when occupants have stopped interacting with their individual workstations and have a higher opportunity for social and/or collaborative interaction in the physical space of a building. In a case study, we benchmark the performance of our proposed method against against socio-organizational network data obtained through traditional surveys. Through statistical tests, we find that our method can uncover network structure that is substantially similar to the survey network. We also find that methods proposed in the literature for network inference through time-series analysis perform less well in this regard. We discuss the subtle characteristics of network inference, pointing to areas in which our method performs well and to areas in which further research will shed more light on organizational structure. We also discuss how building design could be improved to facilitate and promote collaboration among office workers.

3.2 Methodology

We represent a social network as a graph $G = (\mathcal{V}, \mathbf{A})$ where $\mathcal{V}$ is a set of nodes and $\mathbf{A}$ its adjacency matrix. Our objective in this study is to learn the entries of the matrix $\mathbf{A}$ for a given set of nodes (i.e., office workers) within a building. The entry $\mathbf{A}_{i,j}$ is real-valued and represents the strength of the relationship from node $\mathcal{V}_i$ to node $\mathcal{V}_j$. The graph can be undirected, such that any entry $\mathbf{A}_{i,j} = \mathbf{A}_{j,i}$, or directed, such that each value is distinct. The construction of physical networks is typically defined through natural data on the structure of the physical system (e.g., power-line connections between substations in a power grid). In social systems, network data is typically measured through surveys that ask study participants to identify the presence or strength of relationships with the others in the study. Data-driven network inference has relied on analysis of correlations in time series data ascribed to each node (Hallac et al. 2017). This is the perspective we adopt in this study: we aim to learn the adjacency matrix $\mathbf{A}$ of the occupant socio-organizational network by leveraging node-level data streams collected in a building.

Our specific objective for learning $\mathbf{A}$ is to measure behavioral correlations among individuals through time series data. Our methods make use of plug load energy data collected at each workstation—which has been shown to accurately describe occupants' use of space in buildings, thereby offering insight into patterns of behavior that could reveal ties (Sonta et al. 2018; Zhao et al. 2014). We define the time series energy data for occupant i as $\mathbf{X}_i$ (with I as the total number of occupants) and the total number of time steps as $D \cdot T$, where D is the number of days and T is the number of time steps during the day (i.e., if we collect data at 15-minute intervals, $T = 96$). As a preprocessing step, we leverage the method introduced in (Sonta et al. 2018) to map the raw data to abstracted states of occupant activities: $\mathbf{X}_{i,d} \mapsto \mathbf{S}_{i,d}$ (see section 2.3 for details). The mapping uses variational Bayesian inference with a Gaussian Mixture Model to cluster the time series plug load data into states, which can be interpreted as abstractions of occupant activities. In past work, we have typically found three states as the most common number of activity states: low energy, medium energy, and high energy. For example, higher energy use values map to high energy activity states, which correspond to occupants actively using their workstations.

Through a survey of the literature, we identify two existing network learning methods as applicable to this problem statement—the Graphical Lasso and the Influence Model—but we note that specific knowledge of occupant space use creates the opportunity for the design of our own algorithm. Below, we describe these existing methods (section 3.2.1) as well as our own method, which we call the Interaction Model (section 3.2.2). Knowing that human relationships in workspaces are built on interactions, we hypothesize that a new algorithm that directly makes use of occupants' use of space can uncover network structure more accurately than the other methods we implement. To test this approach, we conducted a case study in an 18-person office environment in Berkeley, CA (described in section 3.2.3). Finally, we conduct an analysis leveraging the space syntax methodology to enable comparison of the socio-organizational and spatial structures of the building (section 3.2.4).

3.2.1 Existing data-driven network construction methods

Graphical Lasso: The Graphical Lasso was developed as an algorithm for inferring sparse undirected graphical models—also known as Markov random fields—through L_1 (lasso) regularization. In the literature (Friedman et al. 2008), the data are defined as N multivariate observations in a Gaussian distribution with dimension p, mean μ, and covariance $\mathbf{\Sigma}$. In our case, $N = D{\cdot}T$ (the total number of timesteps), and $p = I$ (the number of workstations/individuals). The Graphical Lasso makes use of coordinate descent to estimate the inverse covariance matrix ($\mathbf{\Sigma}^{-1}$). The covariance matrix can then

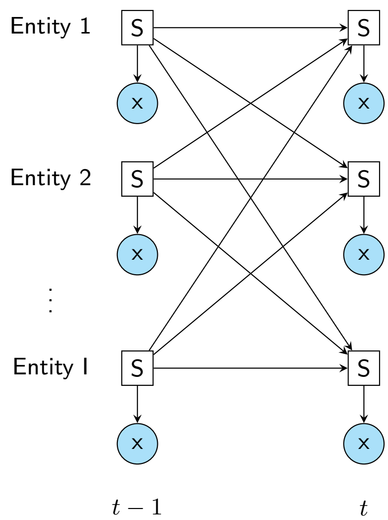

Figure 3.1: Graphical representation of the Influence Model. The model estimates the strength of the arrows between any two time steps using Expectation Maximization, and these strengths form the basis for the inferred network.

be considered as the adjacency matrix in a Markov random field. Because the model assumes that the data are normally distributed, we apply this algorithm to the raw time series energy data, rather than the mapped activity states. We used the R package "glasso" for implementation. Previous work has suggested that this approach can be useful for inferring social network structure from observing the activities of the actors rather than the network itself (Hallac et al. 2017).

Influence Model: The Influence Model, discussed in Pan et al. (2012) models the interaction among entities as a generally coupled Hidden Markov Model (HMM), in which the state of each entity at any given time point is determined by the state of all entities at the previous time point. The model is defined such that the entities—in our case, the building occupants—are in one of a set number of states at any given time point, which ultimately leads them to produce a signal. This has a direct corollary in our problem setting: these abstracted states can be viewed as our occupant activity states and the signals as the plug load energy signatures. A graphical representation of the Influence Model is shown in Figure 3.1. The authors use Expectation Maximization to estimate the parameters of the model by directly using the states, making our mapping from energy to states necessary for this model. One of the key parameters that is learned is the matrix **R**—the "influence matrix"—which can be interpreted as an adjacency matrix in a network. Past work has shown the Influence Model can be used for both physical systems (e.g., a power grid network) as well as social systems (e.g., influence among people conversing) (Pan et al. 2012). We implemented the Influence Model using the MATLAB package available through Pan et al. (2012).

The Graphical Lasso and Influence Model were designed to measure structural relationships among entities, a strategy we argue is best suited for relationships that are distinctly nonsocial (e.g., protein interactions, power grids). The Graphical Lasso can be seen as purely data-driven, directly estimating the network through raw data. The Influence Model makes use of abstracted states to which we map the energy values, but there exists still an opportunity to leverage further information embedded in these activity states. Specifically, we note that different energy use states suggest different uses of the building. We argue that human relationships in organizations are built on opportunities for social interaction in the context of the spatial arrangement of a building, warranting a new method that leverages localized use of space.

3.2.2 The Interaction Model

The Interaction Model (depicted on example data from a single day in Figure 3.2) begins by abstracting the time series plug load data (Figure 3.2a) into states of activities (Figure 3.2b), as discussed in section 2.3. We note that different activity states suggest different uses of the building, and we interpret these states as localized use of the building's space. A benefit of this localized approach is that the spatial and temporal granularity of the data allows one to ascribe patterns of activities and space use to each individual. Leveraging the information embedded in this use of space, the model identifies times when office workers have opportunities for social interaction with other office workers in the building (Figure 3.2c). These opportunities are aggregated using the Jaccard index to form the inferred socio-organizational network (Figure 3.2d).

We assume that in a high energy activity state, occupants are likely to be fully utilizing their equipment at their workstations and are less likely to be moving around. At a medium or low energy activity state, the equipment has entered a power-saving mode or turned off altogether, which in a modern workspace we assume indicates that occupants are more likely to be away from their workstations with higher opportunity for interacting with the people around them. These assumptions are based on our previous validation work (Sonta et al. 2018). We leverage this detail in the data by defining an *opportunity for social interaction* between two occupants to occur when those two occupants have both transitioned from the high energy activity state to either a medium or low energy activity state. We limit these opportunities for interaction such that they only occur after the first transition to the high energy state (i.e., the occupant first arriving at their workstation) and before the last transition to the low energy state (i.e., the occupant leaving for the day). Between these two temporal bounds that bookend the workday, we count the number of times throughout the day that any two occupants have the opportunity to interact. This counting of overlaps in opportunities for interactions forms the weightings of ties among occupants. Formally, we weight the entry $\mathbf{A}_{i,j}$ (the tie strength between occupant i and occupant j) as the Jaccard similarity coefficient of the overlap in opportunities for interaction between occupants i and j. This network inference definition creates an undirected network. We note that previous work on estimating social

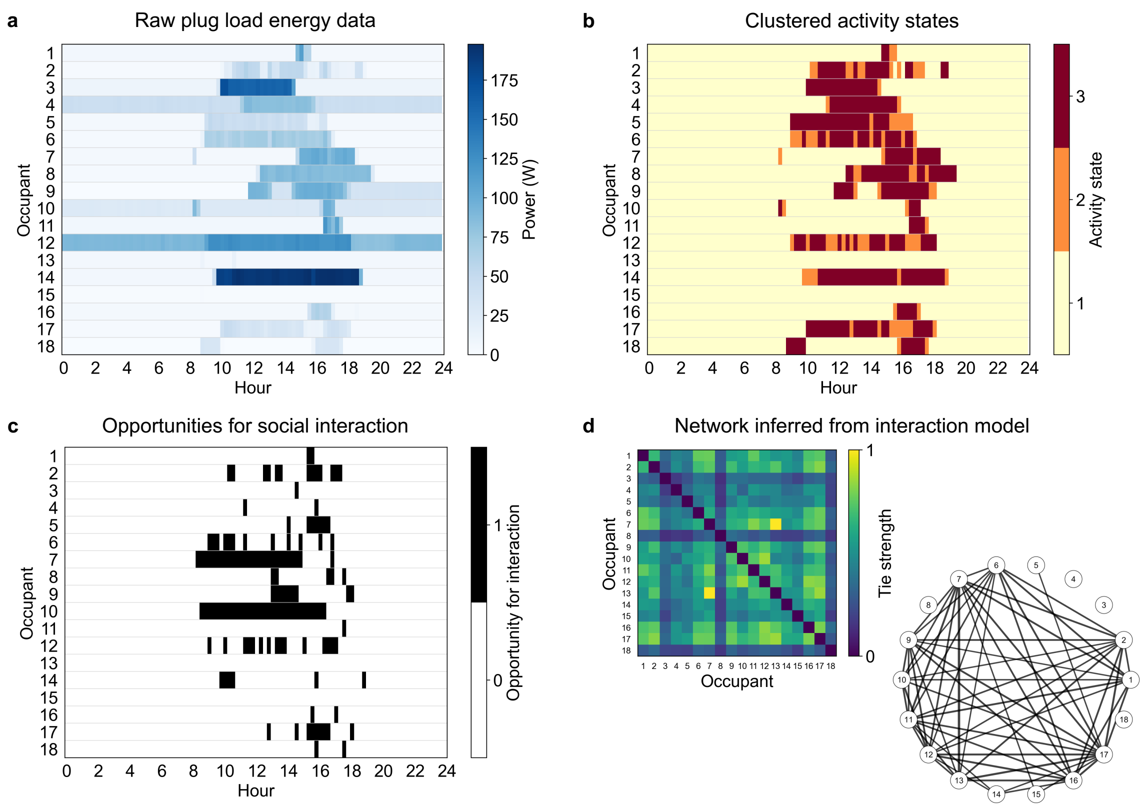

Figure 3.2: Demonstration of Interaction Model steps applied to data from a single day. (a) Raw plug load energy values collected through sensors. (b) Activity states (low, medium, and high energy) resulting from the clustering of the energy data using a variational Bayesian Gaussian Mixture Model. (c) Opportunities for social interaction, where a black value of 1 indicates interaction opportunity. (d) Resulting network (after 365 days of analysis), shown both as an adjacency matrix heatmap and as a graph visualization.

networks on the urban scale has also made use of the general notion of "interaction opportunity" as defined by physical co-presence in an urban setting (Y. Shen et al. 2019).

Given a time series of activity states $\mathbf{S}_{i,d}$ (as described above), the objective of the Interaction Model is to construct the adjacency matrix $\mathbf{A}$ such that the entry $\mathbf{A}_{i,j}$ defines the relationship from occupant i to occupant j. We define the opportunity for interaction as undirected rather than directed, and therefore $\mathbf{A}_{i,j} = \mathbf{A}_{j,i}$. Our algorithm for the Interaction Model is described here (see algorithm 1 for psuedocode). The algorithm first computes vectors for each occupant for each day describing when that occupant has an opportunity for social interaction. This opportunity is defined as being when the occupant is in a medium energy state or when the occupant is in a low energy state between high or medium energy states (i.e., not the before the start of the workday or after the end of the workday). We therefore define the vector $\mathbf{V}_{i,d}$ (for occupant i on day d) as a series of integers of length T, the number of time steps in a day. Each integer is either a zero or a one, with ones indicating the occupant has the opportunity for social interaction; zeros indicating no opportunity. For each day and each pair of occupants, we compute the Jaccard index between the two vectors $\mathbf{V}_{i,d}$ and $\mathbf{V}_{j,d}$. This computation is repeated for each pair of occupants within a day, creating an adjacency matrix $\mathbf{A}_d$ associated with each day. This process is repeated for all days in the study, and the overall adjacency matrix is the average of all $\mathbf{A}_1, ..., \mathbf{A}_D$. The Jaccard similarity between any two vectors $\mathbf{V}_1$ and $\mathbf{V}_2$ is computed as follows:

$$\text{Jaccard}(\mathbf{V}_1, \mathbf{V}_2) = \frac{|\mathbf{V}_1 \cap \mathbf{V}_2|}{|\mathbf{V}_1 \cup \mathbf{V}_2|} = \frac{|\mathbf{V}_1 \cap \mathbf{V}_2|}{|\mathbf{V}_1| + |\mathbf{V}_2| - |\mathbf{V}_1 \cap \mathbf{V}_2|} \tag{3.1}$$

We define the size of a vector and the intersection between our vectors of social opportunity as follows:

$$|\mathbf{V}| = \sum_i \mathbf{V}^i \tag{3.2}$$

$$|\mathbf{V}_1 \cap \mathbf{V}_2| = \sum_i \mathbf{V}_1^i \cdot \mathbf{V}_2^i \tag{3.3}$$

Algorithm 1 Interaction model algorithm

Input: Data matrix $\mathbf{S}$ containing occupants' activity states
Output: Adjacency matrix $\mathbf{A}$ defining the occupant network
Given: D = number of days, I = number of occupants, T = number of timesteps in a day

for $d = 0$ to D **do** ▷ for all days
  for $i = 0$ to I **do** ▷ for all occupants
    if $\max(\mathbf{S}_{i,d}) == \min(\mathbf{S}_{i,d})$ **then**
      break ▷ no occupant activity
    start $\leftarrow \text{argmin}(\mathbf{S}_{i,d})[0]$ ▷ first occurrence of low energy state
    end $\leftarrow \text{argmin}(\mathbf{S}_{i,d})[-1]$ ▷ last occurrence of low energy state
    $V_{i,d} \leftarrow$ vector of zeros of length T
    for t = start to end **do**
      if $\mathbf{S}_{i,d}^t \neq \max(\mathbf{S}_{i,d})$ and $\mathbf{S}_{i,d}^t \neq \min(\mathbf{S}_{i,d})$ **then**
        $V_{i,d}^t = 1$ ▷ occupant in medium energy state
    first_med $\leftarrow \text{argmax}(V_{i,d})[0]$ ▷ first occurrence of medium energy state
    last_med $\leftarrow \text{argmax}(V_{i,d})[-1]$ ▷ last occurrence of medium energy state
    for t = first_med to last_med **do**
      if $\mathbf{S}_{i,d}^t == 0$ **then**
        $V_{i,d}^t = 1$ ▷ occupant in low energy state in middle of day
  for $V_{i,d} \in V_d$ **do**
    for $V_{j,d} \in V_d$ **do**
      $\mathbf{A}_{i,j} \leftarrow \mathbf{A}_{i,j}$ + Jaccard($V_{i,d}$, $V_{j,d}$)
$\mathbf{A} \leftarrow \mathbf{A}/D$ ▷ average Jaccard coefficient

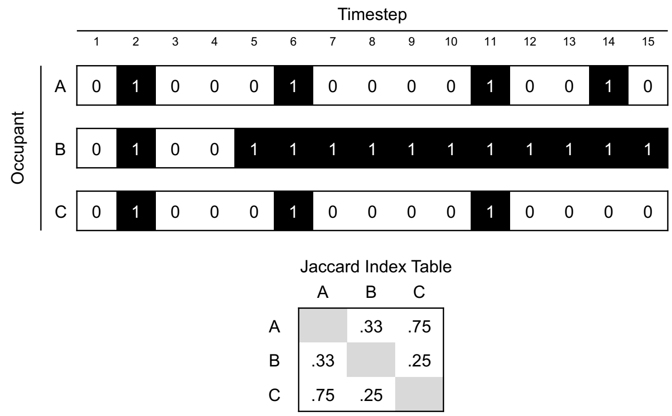

Figure 3.3: Illustrative example showing the normalizing effect of the Jaccard index computed on pairs of vectors describing opportunities for interaction.

The benefit of using the Jaccard index instead of just simply counting the number of overlaps in opportunities for social interaction is that it discounts the situation when a single occupant is in a low or medium energy state throughout much of a workday. In essence, it normalizes these overlaps in opportunities for interaction based on the total number of opportunities that each occupant has in each day. An illustrative example is shown in Figure 3.3.

3.2.3 Data collection and statistical analysis

We installed plug load energy sensors at 18 workstations in the Berkeley, CA office. The office houses a small environmental consulting and engineering firm. Each occupant is assigned to his or her own workstation, allowing us to adopt the perspective that data ascribed to a workstation are ascribed to an individual. The sensors are Zooz smart plugs, which communicate with Z-Wave technology to a Samsung SmartThings hub. Data were collected over 1 year (365 days) between August 2018 and August 2019. Plug load energy data were reported at 1-minute intervals. We aggregate the 1-minute data to 15 minutes, as this time scale has been shown to appropriately describe activities that impact building energy consumption as well as offer insight into social interaction (Sonta et al. 2017, 2018). Manual inspection of the data revealed that plug load energy collection at a very rapid time scale (e.g., less than 5 minutes) can increase noise in the data to detrimental levels, while previous research has shown that a sparse time scale (e.g., greater than 30 minutes) can reduce the amount of valuable information in the data (Melfi et al. 2011). Workstations typically included desktop computers and monitors, as well as some miscellaneous equipment (e.g., phone chargers, lamps, etc.). We made no alterations to the settings of the computers, such as when computers are set to sleep. This decision to minimize interventions was made to maximize extensibility of the overarching methodology to new settings.

The sensors occasionally report erroneous values or fail to report values. As a data cleansing step, we remove data points greater than three standard deviations away from each occupant's mean (0.002% of data points), those less than 0 W (0.0001% of data points) and those greater than 200 W (0.4% of data points). Manual inspection of the data showed that values outside this range are most likely instances of data point corruption during the data collection process. The standard reason for missing data is that the power consumption has not varied—this is a characteristic of the Zooz sensors. This is most common when the power consumption is exactly 0, as any amount of power consumption above 0 typically has natural variation. Because data are reported at 1-minute intervals and we aggregate to 15-minute time steps, we fill forward missing values up to a limit of 15 minutes. This strategy is based on the assumption that these missing values really indicate a power consumption value of 0, but it also allows for the rare case that the power consumption is a steady positive value. If further data points are missing, it is most likely because of connectivity issues or a long period without power consumption. We fill these data points with values of 0.

We collected social and organizational network data through a survey. The occupants of the building were first asked to report with whom in the building they have either a social or organizational relationship. They survey then moves on to two sections, a social weighting and an organizational weighting. The social section asks occupants to rate on a Likert scale (1–7) how closely they view their social relationship in terms of the "inclusion of the other in the self" scale (Gächter et al. 2015; Misra et al. 2016) (see Appendix B). The organizational component asks occupants to mark with whom they (a) share information with, (b) seek technical advice from, and (c) seek personal work-related advice from. These questions measure communication, advice, and trust, as discussed in seminal work by Krackhardt and Hanson (1993). The survey response rate was 83%. The survey forms a directed network—unlike the Interaction Model—but we note that the graph comparison tools we discuss below are valid for comparisons between directed and undirected networks. While interactions are best described without direction, individual views of relationships are inherently directed, and we note that this may be a limitation of the Interaction Model. Details of the survey are discussed in Appendix B.

An important component of the study is to determine the extent to which inferred socio-organizational network structures correlate with the survey socio-organizational structure. Because we define the networks in terms of an adjacency matrix, metrics that measure the correlation between two square matrices can be employed. It is well-known in network analysis theory that standard ordinary least squares (OLS) models are unreliable when autocorrelation is present in the data, as we expect in social network analysis (e.g., within-row and within-column data can be expected to be autocorrelated, since they refer to the same individuals' survey responses) (Krackhardt 1987). Therefore, we adopt the quadratic assignment procedure (QAP) as a permutation test to determine the significance of correlations between two adjacency matrices. We use the Pearson product-moment correlation coefficient to measure the correlation between two matrices, with explicit disregard for the diagonals of the matrices, which we expect to all be zero, since self-loops are disregarded. This combination of the QAP permutation test and the Pearson product-moment correlation has been demonstrated in previous social network research (Henry 2011). The correlation for two graphs G and H is measured as follows:

$$cor(G,H) = \frac{cov(G,H)}{\sqrt{cov(G,G) \cdot cov(H,H)}} \tag{3.4}$$

where

$$cov(G,H) = \frac{1}{|V|^2 - |V|} \sum_{i \neq j} (\mathbf{A}_{ij}^G - \mu_G)(\mathbf{A}_{ij}^H - \mu_H) \tag{3.5}$$

such that i is not equal to j, where V is the set of nodes in the graph, $\mathbf{A}$ is the adjacency matrix, and μ is the mean edge value in the graph. The QAP permutes both the rows and columns of one of the adjacency matrices—a total of $|V|!$ possible mappings of vertices to the edges in the network. The correlation metric can be calculated between the base matrix and all possible permutations of the second matrix, creating a distribution of correlation metrics. By determining where the original correlation falls on the permuted correlation distribution, one can estimate the significance of the correlation (i.e., what fraction of correlations fall below the calculated correlation). In our results discussed above, $|V|$ is 18, causing the number of samples for the correlation statistic to be 306 (i.e., the number of possible ties among individuals, excluding self-loops).

Over the course of the year, three changes happened to the organization occupying the office building, which we were made aware of through direct communication with the organization. After 61 days of data collection, one of the occupants left the organization temporarily. After 258 days, another occupant left the organization indefinitely. After 312 days, the occupant that first left the organization returned, and another third occupant left the organization indefinitely. We adapted the models we employ for network inference to account for these changes. For the Influence Model, we adjusted its mechanics to normalize for the amount of time that an occupant is part of the organization. For the occupants that were part of the organization for a fraction of the year, opportunities for interaction between that occupant and all other occupants were only considered during the times that these occupants were present. In effect, we compute a weighted average of all Jaccard similarities based on the amount of time two occupants were both a part of the organization. This reshuffling created four distinct periods of data collection and analysis that are all very similar but have slightly different organizational makeup.

For example, if all occupants are considered members during the days contained in a vector $\mathbf{d}_1$ and all but one (occupant x) are considered members during the days in $\mathbf{d}_2$, we compute two different adjacency matrices:

$$\mathbf{A}_1 = \text{Interaction}(\mathbf{S}_{d \in \mathbf{d}_1}) \tag{3.6}$$

$$\mathbf{A}_2 = \text{Interaction}(\mathbf{S}_{d \in \mathbf{d}_2}) \tag{3.7}$$

Next, instead of simply averaging all A matrices, we compute the overall adjacency matrix as follows, based on the index of occupant x:

$$\mathbf{A} = \frac{\mathbf{A}_1 + \mathbf{A}_2}{|\mathbf{d}_1| \ + \ |\mathbf{d}_2|} \quad \forall i \neq x, \forall j \neq x \tag{3.8}$$

$$\mathbf{A} = \frac{\mathbf{A}_1}{|\mathbf{d}_1|} \quad \forall i = x, \forall j = x \tag{3.9}$$

3.2.4 Space syntax

The theory of space syntax (Bafna 2003) suggests that the ordering of space impacts how people use space and interact with one another. Recently, researchers have suggested that spatial relationships are correlated with social behavior, such as collaborations in academic settings and “contacts” in workplaces (Potter et al. 2015; Sailer & McCulloh 2012; Wineman et al. 2009). Using a space syntax technique known as axial line decomposition, we constructed networks of the office space in Berkeley, CA, describing both topological distance and angular distance. The topological distance can be interpreted as the number of individual spaces that need to be traversed between workstations, and the angular distance can be interpreted as the amount of physical rotation required to reach a workstation. The space syntax procedure begins with graphically representing the physical barriers within a building in a floorplan. We adopted the axial line decomposition procedure for the spaces that include a workstation in our study. This axial line decomposition procedure begins with the drawing of straight lines that connect all relevant spaces within a building’s floorplan. We note that in certain floorplans with limited barriers, such as long open plan offices, the axial line decomposition technique may have limited ability to explain connections between spaces, though previous research has applied this technique to open plan offices (Sailer & McCulloh 2012). The topological distance between two workstations is defined as the number of line segments that are traversed by traveling from one workstation to another, where a line segment is created whenever two of these lines intersect. The angular distance is defined as the amount of changing of direction that must take place to travel from one desk to another along these lines, where a 90 turn is counted as a value of 1. The specific axial line decomposition for the office building in Berkeley, CA used in this study is shown in Figure 3.4. Leveraging the QAP test described above, we can compare these spatial networks to the inferred and survey-based socio-organizational networks.

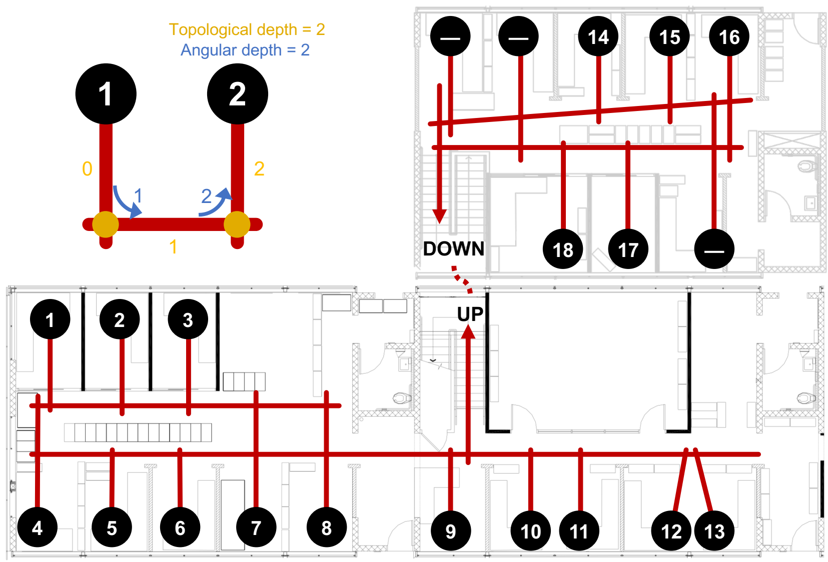

Figure 3.4: Axial line decomposition using space syntax methodology. The floorplan shows the physical spatial barriers, red lines show the line segments connecting individual spaces. The detailed example for workstations 1 and 2 shows the calculation of topological and angular depths.

3.3 Results

3.3.1 Performance of Interaction Model

Figure 3.5 shows the calculated correlations between the combined survey network ($\mathbf{A}^{\text{survey}}$) and each of the three networks inferred with the Graphical Lasso ($\mathbf{A}^{\text{glasso}}$), the Influence Model ($\mathbf{A}^{\text{influence}}$), and our proposed Interaction Model ($\mathbf{A}^{\text{interaction}}$),where the correlation is shown on the reconstructed distribution as calculated through the QAP permutation test. For even a moderately sized graph with 18 vertices, computing the full distribution of graph correlations through permutation is computationally prohibitive; we therefore employ Monte Carlo simulation with 10,000 repetitions to estimate the distribution. We can see from Figure 3.5 that only the Interaction Model produces a significant correlation at 0.46 with an estimated p-value from the permutation test of 0.002. Figure 3.6 shows the overall correlation matrix (upper half) between each of the three learned networks and each component of the survey network, including the combined organizational survey network and the overall survey network. Each entry in the matrix shows the correlation between the graphs denoted along the diagonal. We note that the Interaction Model has similar correlations with each component of the survey network, with no specific component having much larger or much smaller correlations with the Interaction Model than the overall survey network.

The correlations and estimated significance levels (Figure 3.5 and Figure 3.6) offer strong supporting evidence for the relative success of the Interaction Model in capturing office worker relationships

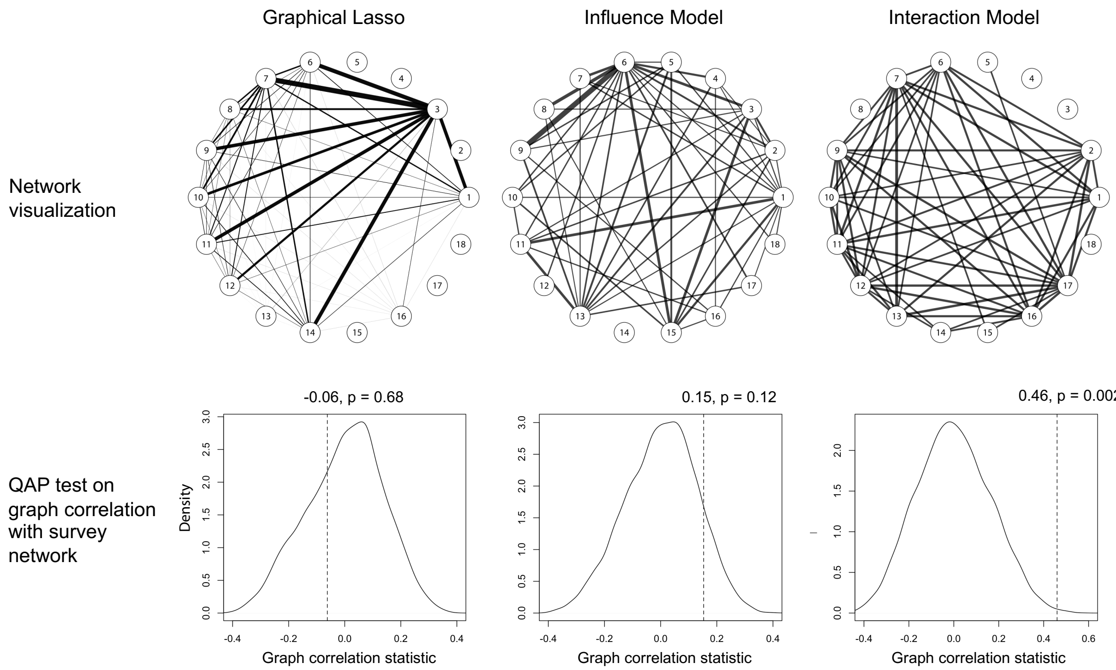

Figure 3.5: Inferred network visualization and statistical results. Network visualization for each of the three network inference methods, along with Pearson product-moment correlation with the survey network over the distribution from the Quadratic Assignment Procedure (QAP) test. Vertical lines on the distributions indicate the measured correlations. Comparing these measured correlations to the distribution enables the estimation of the p-values for these correlations (Graphical Lasso: p=0.68; Influence Model: p=0.12; Interaction Model: p=0.002). Note: A threshold is applied to tie strengths in each graph for visualization purposes.

as compared to the Graphical Lasso and the Influence Model. The overall survey network is based on social and organizational weights, where the social weighting is based on one survey question and the organizational weighted is based on three. There are therefore four individual components of the overall survey network. While the Interaction Model has relatively similar correlations with all of these individual components, the correlations are highest for the trust component of the organizational survey network and for the social survey network. Among the organizational components of the survey network, the communication and advice networks have lower correlations, perhaps due to the fact that these networks have higher graph densities (0.70 and 0.54, respectively) compared with the trust graph (0.42). In other words, each occupant communicates with and seeks advice from most of his or her colleagues, while seeking trust is more selective. The more fully connected communication and advice networks would not be expected to correlate as well with the inferred networks, because there is less differentiation and social choice among their ties.

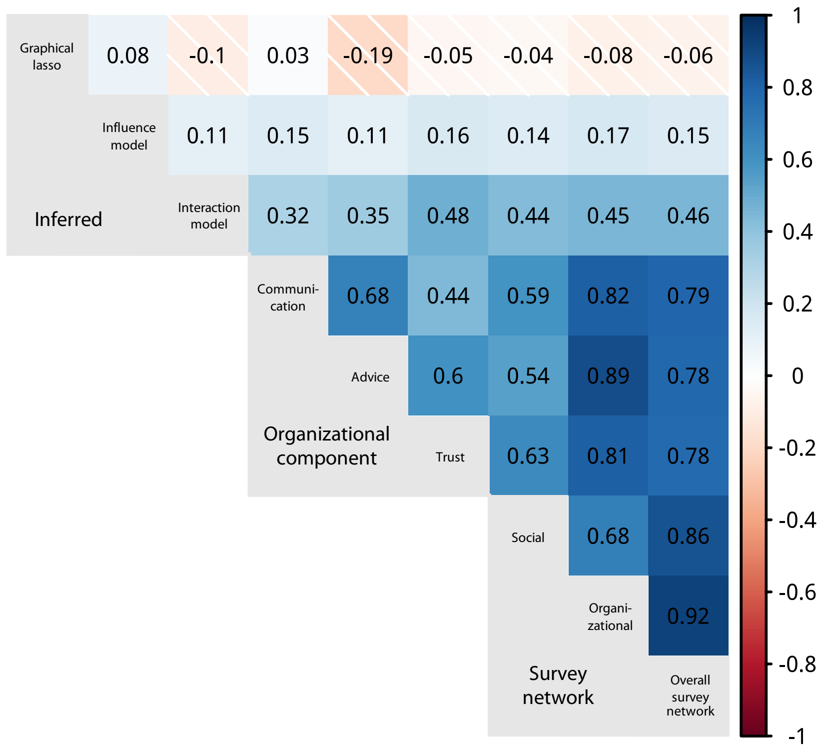

Figure 3.6: Correlations among inferred and survey networks. Correlogram showing Pearson product-moment correlations between the three inferred networks, each component of the organizational survey network, and the social and organizational survey networks.

3.3.2 Temporal analysis of Interaction Model

One strength of the Interaction Model is that its key building blocks—opportunities for social interaction—are assigned to individual time steps, allowing our analysis to capture the temporal dimension of office worker interactions. The model estimates times at which any two occupants have stepped away from their workstations and defines these instances as opportunities for social interaction. Figure 3.7a shows the distribution of total overlaps in social opportunity aggregated across all days for the average occupant, plus or minus one standard deviation calculated across occupants. We can see from this figure that there is a large spike between 12 pm and 2 pm, which is likely to be the standard time for lunch in this particular office. Figure 3.7b shows these opportunities for interaction by day, where we see a similar lunchtime spike for each day. However, there is also a large peak on Monday near 10 am. Based on knowledge of the operations of this organization, we note that there is a large meeting involving all members of the organization at this time. This serves as observational validation that the Interaction Model is in fact capturing opportunities for interaction.

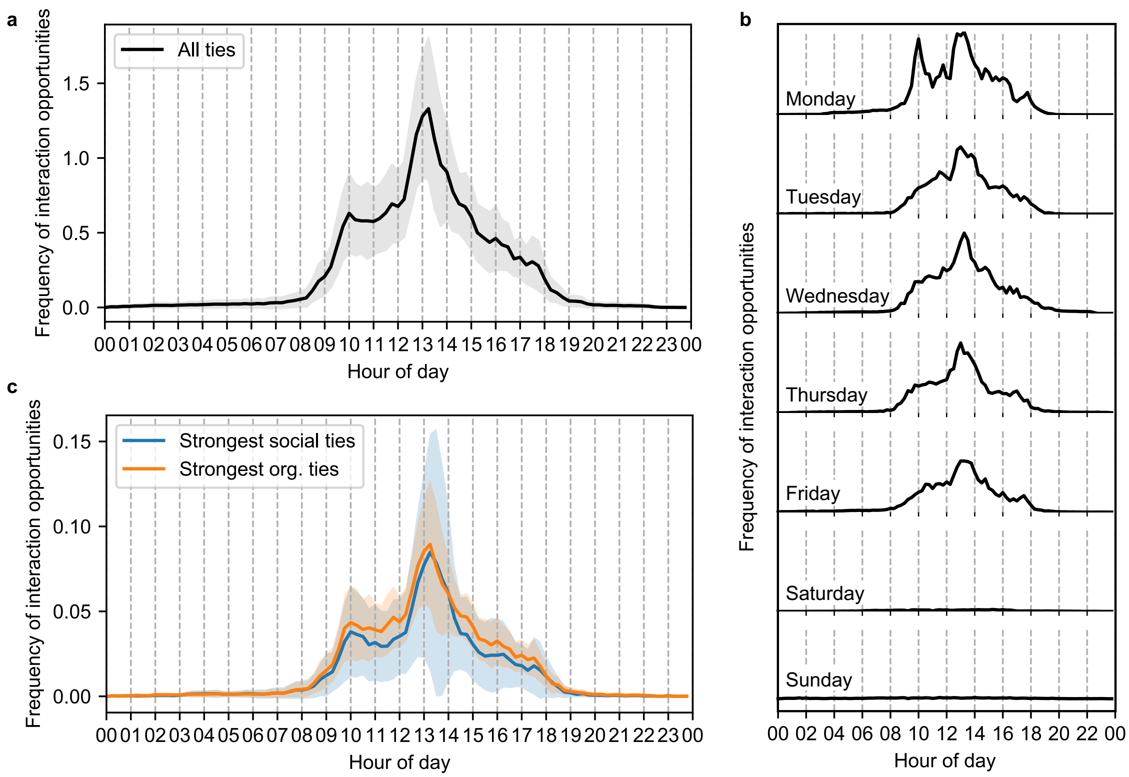

Figure 3.7: Frequency of opportunities for social interaction over 24 hours, aggregated over data collection period. Lines represent the average across individuals and shading shows one standard deviation across individuals. (a) All opportunities for interaction for all ties. (b) Comparison of interaction opportunities by day of the week. (c) Comparison of interaction opportunities for each occupant's strongest social ties and strongest organizational ties, as defined by the survey network.

Because the survey network is comprised of both social and organizational ties, a natural question is to ask whether one can differentiate between these ties through more detailed analysis. We therefore sought to characterize, leveraging the survey results, whether social ties or organizational ties are more likely to manifest themselves at different times of the day (e.g., Do people eat lunch with their friends? Do they meet with their work ties in the morning?). Figure 3.7c shows the temporal distribution of opportunities for social interaction of two different kinds of true ties defined by the survey network—the strongest social tie for each occupant and the strong organizational tie for each occupant. This breakdown between ties shows that overall trends are very similar for these two relationships, but we note two points of interest. First, the standard deviation of interaction opportunities is higher for social ties than for organizational ties around lunch time, suggesting that there are some pairs of occupants with social ties that have many interactions around lunch time. Second, there seems to be a larger count of interaction opportunities in the morning before noon for organizational ties, though the increase over the social ties is within the standard deviations.

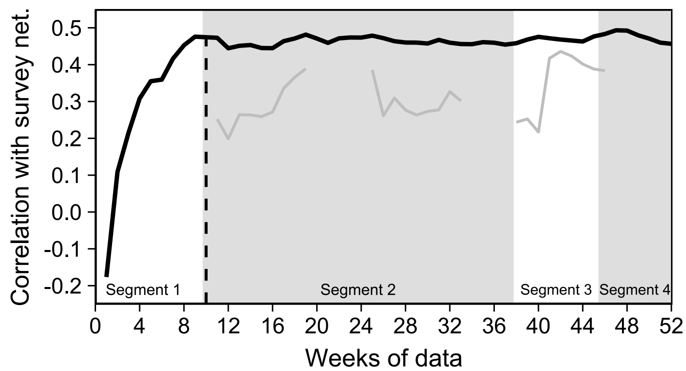

Figure 3.8: Graph correlation over time. Each shift in data segments refers to a small change in the organizational structure (i.e., an occupant leaving the organization or returning to the organization).

3.3.3 How much data is needed for the Interaction Model?

One question that any inference built on time series raises is the length of time required to build a successful model. In order to test this, we created 52 batches of the data based on timestamps (i.e., batch 1 contained the first week of the data, batch 2 the second week, etc.) We first applied the Interaction Model to growing amounts of the data (i.e., the first week, then the first two weeks, and so on) and calculated the correlation between that network and the survey network. Figure 3.8 shows how the graph correlation varies over time as the data grows (black line). As a result of three minor organizational changes (i.e., occupants temporarily or permanently leaving the organization), there are four distinct data segments (see Materials and Methods). After roughly 10 weeks (20% of the year), the increase in correlation experiences significant diminishing returns.

Each of the points on the black line in Figure 3.8 use a starting point of week 1. To investigate the impact of when data collection begins, we also applied the Interaction Model to 10 weeks of data (the threshold determined above) using different starting times (gray lines). These lines also show the evolution of the correlation as the data grows (e.g., for the furthest left gray line, week 10, then weeks 10 and 11, and so on). None of the three starting times produced a graph correlation as high as the original starting time, indicating that true social ties may have changed as a result of organizational changes or other time effects.

3.3.4 Relating spatial and organizational networks

Figure 3.9 shows the Pearson product-moment correlations among the topological spatial network, the angular spatial network, the network inferred through the Interaction Model, and the overall survey network. The correlations between the space syntax networks and both the Interaction Model and the survey network are negative. We would expect this negative correlation, since a larger distance between occupants' workstations has been shown to inhibit communication and interaction

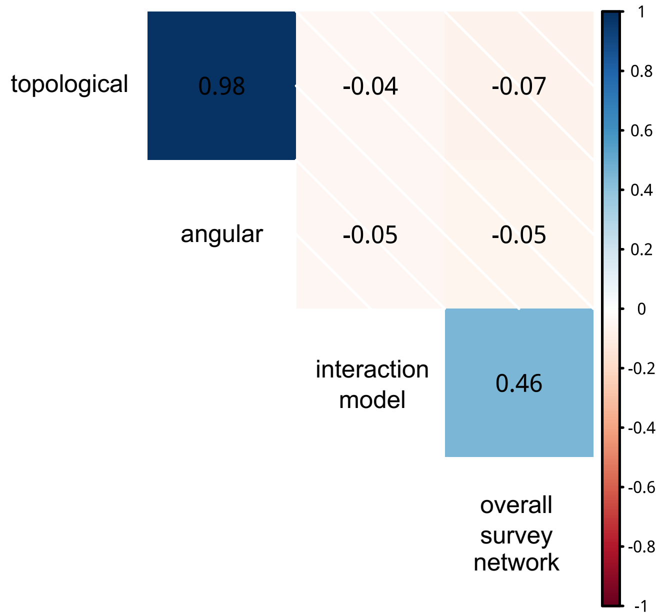

Figure 3.9: Correlations between spatial, inferred, and survey networks. The topological and angular networks, inferred through two space syntax methods, describe the spatial relationships between each pair of workstations. The interaction model and survey networks describe socio-organizational relationships for the individuals associated with each workstation.

(Sailer & McCulloh 2012). However, these correlations are all insignificant. The lack of a strong correlation between the spatial layout of the building and the interactions among occupants as well as the true socio-organizational relationships is surprising but suggests that there is an opportunity to design the layout of the building in such a way that actually promotes collaboration. For example, if a new layout were chosen such that the distances between desks was perfectly negatively correlated with the true organizational structure, the new layout might promote more meaningful interactions in space.

3.4 Discussion

Our results suggest that it is possible to capture a significant component of the true socio-organizational structure of organizations in buildings by analyzing sensors signaling occupant behavior. Methods adapted from the literature—the Graphical Lasso and the Influence Model—were limited in their abilities to reconstruct the true socio-organizational structure of office workers. Each of these methods may suffer from limitations in the model construction. The Graphical Lasso was designed to measure correlations in raw time series. However, the noise in the raw plug load energy data warranted the need to abstract the data into states of activities in order to enable analysis of true patterns of space use, and therefore the noise may have impacted the Graphical Lasso's efficacy. The Influence Model does explicitly incorporate latent states in its modeling approach (in our case, activity states), but it was originally designed for non-social systems (e.g., power grids) where interactions can be defined by physical laws. However, manifestations of relationships among people can be characterized as impromptu and unstructured, which are details we sought to address in our modeling approach. Our proposed method—the Interaction Model—performed significantly well. We theorize the model is successful because it is built on the knowledge gained by analyzing occupants' uses of their individual spaces. The underlying data from plug load sensors installed at the desk level, when abstracted to states of occupants' activities, offers rich insight into use of space. These insights enable the analysis of opportunities for social interaction that are the foundation of our network inference strategy. We note, however, that the specific strategies implemented in our research could be extended to use other sources of raw data that can be used to model occupancy, such as infrared or lidar sensors.

We demonstrate in a case study that about 10 weeks of data collection was sufficient to create a network that is significantly similar to the network obtained through traditional survey instruments. After 10 weeks, the model experienced substantial diminishing returns. This behavior could arise from a few different reasons. One set of possibilities concerns the fact that we distributed the survey very near the time that we started collecting data. It is possible that the true network changed after the first organizational change (shift from data segment 1 to data segment 2 in Figure 3.8), and therefore the survey network no longer represented the true network. In this case, any additional information collected through the data would be biased toward the original network structure as represented in the survey network. It is also possible that the network changed continuously over the course of the year, and as a result, occupants' behavior changed as well. In this case, the network inferred with enough data closest to the time that people reported ties would be most accurate. On the other hand, it is possible that the act of distributing the survey caused occupants to become more conscious of their social behavior and therefore interact more with the people they reported ties with. This kind of behavior change as a result of an external trigger—the "mere-measurement effect"—has been documented in previous social network studies (Sprott et al. 2006). Our additional analysis of running the Interaction Model using different start dates shows that using later start dates reduces the correlation between the survey network and the Interaction Model network. This finding aligns with the notion that the inferred network correlates best with the survey network when the data is analyzed closest to the time that the survey is distributed.

While the Interaction Model did experience diminishing returns, it was able to reconstruct a significant component of the survey network. This success points to substantial opportunities to leverage low-cost, ubiquitous ambient sensing technologies for the study of organizational networks present in office spaces. Researchers studying social networks and organizational design have noted the importance of understanding network structures due to their impact on organizational outcomes related to the broad notion of productivity, such as creativity and collaboration (Perry-Smith & Mannucci 2017; Uzzi & Spiro 2005). Our findings demonstrate that socio-organizational networks can effectively be mined from data. As a result, the potential for network analysis in organizations can be extended to situations previously not possible due to the cost and time of survey-based network construction, such as large organizations occupying large buildings or building campuses. Abstractions of the ambient sensing data will also enable further analysis of network effects using inherently anonymous data that does not contain personally identifiable information beyond the desk location. While the desk location could be traced to an individual if seat assignments are known, the underlying power consumption data is an inherent level of abstraction away from individual actions. This ambient approach to modeling of human systems, as opposed to direct observation of actions, is an important aspect given new laws and protocols surrounding data privacy (e.g., GPDR (*Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (GDPR)* 2016)).

We additionally show that in the context of the space analyzed in our case study, the spatial layout was not significantly correlated with the true structure of the organization, suggesting that a redesign of the layout could enhance interaction and improve collaboration among the members of the organization. Researchers studying the relationship between space and organizations have linked space to the nexus between networks and organizational outcomes, typically showing that spatial proximity increases the likelihood for relational tie formation and the positive outcomes associated with these ties (Sailer & McCulloh 2012). The tools presented in this paper offer an opportunity to test these previously formulated hypothesis at further scale. Additionally, when it comes to re-designing the spatial layout of organizations, the ability to quickly and continuously measure the social network—as our research enables—will allow network structure to become an attainable and important input into this design process. Future work can incorporate network structure into the design of space itself, as engineers and designers continue researching methods for optimizing spatial layouts through generative design (Nagy et al. 2017). Moreover, as spaces become more flexible, the ability to capture a social network through analysis of data collected over a period of 10 weeks will create the opportunity to dynamically design building layouts in an effort to enhance collaboration. This dynamic model of building and space design will become increasingly important given the rise of new office space procurement models such as co-working spaces and shared common areas.

One important limitation of the Interaction Model concerns the notion of opportunities for social interaction. There are many occasions that might result in two occupants being in a state of possible social interaction as defined above, such as two occupants taking separate trips to different break rooms. While some of these occasions could be unrelated to meaningful interactions, we assume that repeated instances of mutual opportunity for interaction would outweigh spurious instances of similar activity patterns. In our analysis of the amount of data required for our model to be successful, we see that the performance of the Interaction Model steadily improves over the first 10 weeks, which supports the notion of the importance of repeated interactions. In a larger building, this limitation could also be addressed by including a spatial weighting that discounts possible interactions that would be physically infeasible. Another limitation concerns the notion of distributing a survey as a means of measuring the "ground truth" as closely as possible. We note that while research has demonstrated the shortcomings of surveys particularly with regard to the problems of self-reporting information (Marsden 2012), surveys remain the standard method for measuring social network structure. Finally, we note that the Interaction Model produces an undirected network, while the individual responses from the survey are directed because they represent each occupant's individual perceptions of the relationship. While physical interactions can best be described in an undirected manner, certain aspects of social relationships (e.g., individual perceptions of friendships, amount of advice given vs. received) can be viewed as directed. We note, however, than many tools common in social network analysis, such as community detection or graph clustering, are traditionally applied to undirected networks (Malliaros & Vazirgiannis 2013). Therefore, the information embedded in the undirected network retains much of the needed information for decision-making based on socio-organizational relationships. Future work could extend our proposed Interaction Model to provide more detailed insights into social relationships within an office space and enable more targeted interventions related to the directionality of communication and/or information flows.

Overall, our results show that the Interaction Model is successful in reconstructing a significant component of the true socio-organizational network. Inferring this network can be useful for analyzing the design of the space (e.g., using space syntax) in the context of the socio-organizational system. Information on the structure of the organizations in commercial buildings will enable the design and management of spaces that promote organizational success. In particular, managers of organizations will be able to better understand which occupants and/or relationships are vital to organizational operations. At the same time, managers of building spaces can use the information to make ad hoc changes to building designs, such as removal of barriers or changes to occupant layouts—changes that can be used to better align the social and spatial networks. While it may not be feasible to continuously re-arrange office spaces based on insights garnered from our proposed methods, our work could the basis for more strategic (e.g., quarterly) reviews of how spaces are being used and how occupant layouts are encouraging or limiting social interactions. Overall, our methods aim to enable a new dynamic perspective of organizational decision-making at the socio-organizational-spatial nexus, whereby offices can be critically analyzed and changed to improve performance.

3.5 Conclusion

The Interaction Model introduced in this paper enables passive measurement of network structure using ambient sensors that are growing in popularity in commercial buildings. Through statistical tests, we show that this network inference strategy learns a network with significant correlation to the ties reported through distribution of a survey. In our case study, we find that roughly 10 weeks of data are required to learn the network with the ambient sensing data. This relatively short data collection period can enable dynamic measurement of occupant networks as organizations evolve and spaces change. We also demonstrate that learning this network enables the analysis of the spatial design of buildings in the context of their human systems. In particular, comparing the learned social network to spatial networks as defined by the space syntax literature creates insight into whether the spatial layout matches socio-organizational behavior. In the case study presented in this paper, we found no significant correlation with the spatial networks, demonstrating the opportunity to design the space in a way that better reflects the socio-organizational network. The Interaction Model can therefore become a significant component of a design strategy that promotes human-centered organizational objectives such as collaboration. The model and analyses introduced here will promote the design, engineering, and management of buildings in a way that allows human and built systems to work together.

Chapter 4

Occupant Energy Signal Processing on Graphs ($\mathbf{OESP_G}$)

A computational framework for multidimensional analysis of occupant energy use data in commercial buildings

This chapter is adapted from the following paper: Sonta, A. J., Jain, R. K., Gulbinas, R., Moura, J. M. F., and Taylor, J. E. (2017). "OESP$_{\mathrm{G}}$: Computational Framework for Multidimensional Analysis of Occupant Energy Use Data in Commercial Buildings." *Journal of Computing in Civil Engineering*, 31(4), 04017017.

Abstract

Commercial buildings account for much of the energy use both in the United States and globally. The role of occupant behavior within the physical building has been found to be an important factor in the overall energy use profile of commercial buildings. Recent research has noted the potential energy savings that can be achieved when occupant behavior is beneficially modified. However, frameworks for analyzing occupant behavior are limited in their ability to simultaneously consider three key dimensions of occupant-driven energy use in buildings: spatial, temporal, and social. In this paper, we present the Occupant Energy Signal Processing on Graphs (OESP$_{\mathrm{G}}$) framework, which is able to address the three key dimensions of occupant energy use in commercial buildings through an inherently scalable mathematical structure. We demonstrate the mechanics, applicability, and merits of the OESP$_{\mathrm{G}}$ framework by applying it to occupant energy use data through both a simulated example and real test-bed data from a commercial office building. We find that OESP$_{\mathrm{G}}$ is able to identify situations in which occupant energy use through plug loads is out-of-sync with what would be expected based on nuanced spatial and organizational identity, and we note the feasibility of using this framework to make recommendations for temporal and spatial occupancy shifts that would have a positive impact on occupant energy use.

4.1 Introduction

Commercial buildings are responsible for nearly 20% of current energy use in the United States, and they are projected to be the fastest growing energy demand sector worldwide (U.S. Energy Information Administration 2016a,b). As a result, researchers are developing new approaches to enhance the energy efficiency of commercial buildings and reduce the associated environmental emissions and negative sustainability impacts of their energy usage. Specifically, new approaches that bridge the gap between physical building systems and the behavior of building occupants have shown significant promise to enhance the energy efficiency of commercial buildings (Azar & Menassa 2012; Hong & Lin 2012; Meier 2006). However, realizing savings from such new approaches will require a comprehensive and nuanced understanding of occupant behavior. This is due to the fact that energy usage has been shown to vary dramatically because of occupant dynamics (Clevenger & Haymaker 2006).

With regard to individual occupants, differences in their locations within a commercial building (*spatial dimension*) as well as variations in their schedules (*temporal dimension*) can have strong implications for how the building uses energy (Jazizadeh et al. 2014; Kwak et al. 2014; Lim et al. 2012). Social and organizational networks of people (*social dimension*) have also been found to be an important factor in the energy use trends within a commercial building (Anderson et al. 2014; Chen et al. 2012; Khashe et al. 2016). As a result, a successful socio-technical and occupant behavior based approach to energy efficiency for commercial buildings will require reconciling and analyzing these three key dimensions (spatial, temporal, social) by which building systems and occupants consume energy.

The advent of data streams from low-cost building sensors offers an opportunity to improve the granularity by which we understand building energy use trends in the temporal, spatial, and social dimensions. This improved understanding from high-fidelity data streams enables optimization of building design and controls so that buildings can run more efficiently. However, the methods typically used to analyze occupant energy use data are limited due to their ability to capture only one or two of the three aforementioned key dimensions affecting energy use in commercial buildings. In this paper, we introduce the Occupant Energy Signal Processing on Graphs (OESP_G) framework, a scalable computational framework for multidimensional analysis of building energy usage data inspired by the emerging field of signal processing on graphs (Sandryhaila & Moura 2013). We demonstrate the mechanics, applicability, and merits of our proposed framework using a simple simulated example as well as a case-study example comprised of real data from a test-bed office building in Denver, CO.

4.2 Background

4.2.1 Occupant and data-driven energy efficiency in buildings

Numerous studies have found that occupant behavior plays a major role in both residential (Gill et al. 2010; Guerra Santin et al. 2009; Majcen et al. 2015; Martinaitis et al. 2015) and commercial (Bonte et al. 2014; Hong & Lin 2012; Norford et al. 1994) building energy performance. Specifically, previous work has indicated that in commercial office spaces, total energy consumption can be expressed as a combination of a baseline amount of energy consumption and human-driven energy consumption (Taherian et al. 2010). This human element, though important in a commercial building's energy use, is typically difficult to characterize. As a result, recent studies have developed new simulation methods and approaches to model and study occupant behavior in commercial buildings. Azar and Menassa (2012) proposed an agent-based simulation modeling method to study occupants' energy use characteristics as they change over time. Occupant-based models such as these can also be coupled with whole-building energy simulation tools to develop models that building management systems can leverage to address the complex interactions between occupancy and building performance, and ultimately help find opportunities for improved energy efficiency (Menassa et al. 2013).

While providing valuable insight, simulation driven methods and approaches are limited in their ability to leverage new high-fidelity and spatially-granular energy usage data streams in their analysis. Such energy usage data streams could provide deeper insight on human-driven energy consumption at the spatial level of the individual occupant and at sub-hourly temporal intervals. As a result, recent work has begun to leverage these data streams to improve our understanding of energy usage in the built environment (Agarwal et al. 2010; Kazmi et al. 2014; Menzel et al. 2008; Milenkovic & Amft 2013). However, in such data-driven studies the multi-dimensionality of occupant energy usage optimization is not directly addressed. In the following section, we review the three key dimensions of occupant energy usage and classify existing data-driven methods and frameworks to elucidate a gap within the current body of work in data-driven occupant energy analysis.

4.2.2 Dimensions of occupant energy use

The way in which occupants use energy in commercial buildings can be characterized by three key dimensions: *spatial*, *temporal*, and *social*. The spatial dimension characterizes where within a building an occupant consumes energy and associated services from building systems. As a result, the spatial dimension is typically considered by building designers and operators by implementing HVAC zoning to improve thermal comfort and, in some cases, energy performance (Smith et al. 2012). Similarly, the temporal dimension characterizes the time of the day in which occupants are present, using energy, and requiring services from building systems (i.e., occupant schedules). From a high level, this is captured through rudimentary occupant schedules that are often built into energy simulations to try and match predicted energy use with actual building energy use

(Clevenger & Haymaker 2006; D'Oca & Hong 2015). The social dimension can be characterized by the organizational network dynamics that describe occupant interactions and aspects of occupant behavior in buildings. The impact of the human element is often overlooked due the challenges in modeling occupant behavior, but recent studies have noted that understanding the social dimension is of great importance in respect to minimizing building energy use (Anderson et al. 2014; Gulbinas & Taylor 2014). For example, the inherent social dynamics that describe occupant behavior within a building have been found to have large implications for the effectiveness of energy-use feedback tools that report real-time energy consumption information to occupants (Gulbinas & Taylor 2014). Given the ability to analyze energy use data on a sub-building level using sensors, recent building energy analysis methods and frameworks that recommend efficiency strategies have begun to address these three dimensions, but in many cases only individually. In the following subsections, we briefly discuss current frameworks and classify them in terms of their primary dimension of concern.

Spatial Dimension: One body of work has looked at improving control of building zones in concert with thermal conditions and preferences of occupants. Such work has been found to improve overall indoor thermal comfort and avoid situations where energy is unnecessarily wasted (Jazizadeh et al. 2013; Schoofs et al. 2011). Azar and Menassa (2016) proposed a data-driven framework to analyze occupancy spatially and propose energy saving actions. Additionally, matching occupant preferences with a decentralized control strategy has been found to have significant energy saving potential (Jazizadeh et al. 2014). A spatially driven analysis framework has also been developed for lighting and has yielded energy savings on the order of 50% in case studies (Krioukov et al. 2011).

Temporal Dimension: Recently developed frameworks for improving building energy efficiency consider the scheduling of building activities, and the matching of building schedules with occupancy predictions and/or measurements (Lim et al. 2012; Majumdar et al. 2012, 2016). The synchronization of occupancy predictions with optimized scheduling of meetings has been tested as a strategy for reducing energy consumption in office buildings (Majumdar et al. 2012). Moreover, recent work has also aimed to temporally characterize and predict occupant energy usage in order to identify patterns that could be utilized to formulate energy efficiency strategies for a commercial building (Gulbinas et al. 2015; Khosrowpour et al. 2016).

Social Dimension: Modeling occupant behavior has been shown to improve the understanding of building occupants' changing energy use characteristics over time (Azar & Menassa 2012). Beyond individual occupant behavior, the organizational network dynamics that follow from the social structure of the building have been found to play an important role in improving building energy efficiency (Anderson & Lee 2016; Khashe et al. 2016; Manika et al. 2013; Siero et al. 1996). As interventions are proposed for energy-efficiency purposes in office buildings, the social network structure of the occupants has been shown to be critical in determining and predicting the absolute effectiveness of the intervention strategy (Anderson & Lee 2016). Additionally, the formation of human networks in buildings has also been found to be influenced by the form of the office building, drawing a connection between the human and the spatial dimension of occupant behavior (Sailer & McCulloh 2012). Despite the growing evidence regarding the impact the social dimension can have on occupant energy usage and commercial building operations, no frameworks have been proposed to analyze the social dimension individually or in tandem with other dimensions.

Thus, there is significant opportunity to further understand how human dynamics and spatial and temporal variability of occupant energy consumption within a building can lead to the identification of energy saving opportunities. Previous studies have been limited by their scope in analyzing all three dimensions of occupant-driven energy efficiency, and, as a result, they may not yield the insight into the complex dynamics of building energy use necessary to maximize energy savings associated with new occupant driven approaches. In this paper, we introduce—and test on real data—the Occupant Energy Signal Processing on Graphs ($\mathrm{OESP_G}$) framework, a scalable computational framework that is capable of analyzing all three dimensions of occupant-driven energy efficiency in buildings. The framework aims to provide a method for identifying situations in which energy use in the study building is not in synchronization with what would be expected based on the temporal patterns, spatial layout and organizational network structure of the building's occupants, thereby simultaneously addressing the three key dimensions of occupant energy usage in commercial buildings.

4.3 Methodology

Applying the $\mathrm{OESP_G}$ framework to building energy use data consists of four main steps: (1) gather data describing the building's spatial layout, social structure, and time-series energy use, (2) construct a graph representing the spatial and organizational structure of the building, (3) analyze building energy use data by representing the data as signals, and (4) characterize the energy use data. In this section, we introduce the framework, highlight the underlying mathematical and graph theory concepts from the literature, and demonstrate the mechanics of the framework using a simulated example. We utilize energy use data at the plug load because it provides a good proxy for changes in occupant behavior (see Appendix A for further information and empirical data).

4.3.1 Discrete Signal Processing on Graphs

The emerging field of signal processing on graphs (Sandryhaila & Moura 2013; Shuman et al. 2013), develops methods of analysis of signals supported by graphs. In particular, the Discrete Signal Processing on Graphs ($\mathrm{DSP_G}$) framework in (Sandryhaila & Moura 2013, 2014b) extends concepts from traditional signal processing to data that can be indexed by vertices on graphs. Signals indexed by graphs arise in many situations where data is collected, including measurements from sensor networks (Akyildiz et al. 2002), community preferences (Leicht & Newman 2008), and many others. A central contribution of the work presented in this paper is to expand and adapt the underlying concepts from $\mathrm{DSP_G}$ for the problem of multi-dimensional analyses of building energy use data. As such, our $\mathrm{OESP_G}$ framework adopts the adjacency matrix of the graph structure as its main building block and utilizes a graph Fourier transform to expand a signal into a Fourier basis in the graph spectral domain.

Consistent with previous work (Sandryhaila & Moura 2013), we define the relationships between data elements (i.e., occupants) as a graph $G = (V, \mathbf{A})$, with N occupant nodes, where $\mathcal{V} = \{v_0, ..., v_{N-1}\}$ is a set of occupant nodes and $\mathbf{A}$ is the weighted adjacency matrix of the graph. Each data element is indexed by an occupant node v_n, and each weighting $\mathbf{A}_{n,m}$ of the edge from v_n to v_m describes the directed weighting from the nth node to the mth node. The distinct eigenvalues $\lambda_0, ..., \lambda_N$ of the adjacency matrix $\mathbf{A}$ are the *graph frequencies* and form the *spectrum* of the graph. The eigenvector corresponding to any graph frequency is the *frequency component* corresponding to that frequency.

For each node, power draw values are continuously collected. The power values are defined for each occupant in the set:

$$\mathbb{P} = \{\mathbf{P}_0, ..., \mathbf{P}_{n-1}\} \ \forall \ n \in N \tag{4.1}$$

where $\mathbb{P}$ = set of all power vectors ($\mathbf{P}_n$) for all N occupants, and n = occupant node index. Each occupant's power values are collected in the vectors defined above, and defined as:

$$\mathbf{P}_n = \{p_n^t, ..., p_n^T\} \tag{4.2}$$

where $\mathbf{P}$ = vector of all power draw values for occupant node n, t = time index, and T = total number of periods of data collection. It is important to note that when analysis is being conducted in near-real time parameter T will continue to grow as data is collected and more time periods are added to the power draw vector $\mathbf{P}_n$.

In order to account for variations in typical power draw values for the different occupant workstations, power values are normalized using a running normalization process:

$$\bar{p}_n^t = \frac{p_n^t}{p_n^{t,max}} \tag{4.3}$$

where $\bar{p}_n^t$ is the normalized power draw value at time t for occupant n, and $p_n^{t,max}$ is defined as:

$$p_n^{t,max} = \max(p_n^{t-c}, ..., p_n^t) \tag{4.4}$$

where c = parameter indicating the number of periods over which the current value is normalized. For example, if $c = 12$, and the time step is chosen to be 1 h, the running normalization normalizes each value over the previous 12 h of data.

Finally, each snapshot of normalized plug load power draw becomes an individual graph signal, defined as a map:

$$\bar{p}_n^t(v_n) \mapsto s_n \tag{4.5}$$

where s_n = graph signal coordinate associated with the occupant node v_n. The graph signal can be represented as a vector:

$$\mathbf{s} = [s_0, ..., s_N]^T \in \mathbb{R}^N \tag{4.6}$$

We utilize a Fourier transform to expand the signal into the graph spectral domain. In this initial work, we assume a graph structure with undirected edges, such that $\mathbf{A}_{n,m} = \mathbf{A}_{m,n}$, causing eigendecomposition of $\mathbf{A}$ to be in the real domain. As such, the eigendecomposition is as follows:

$$\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^{-1} \tag{4.7}$$

and the graph Fourier transform of the signal s is:

$$\hat{S} = \mathbf{F}s \tag{4.8}$$

where $\mathbf{F} = \mathbf{V}^{-1}$ is the graph Fourier transform matrix. The values of $\hat{s}_n$ characterize the *frequency content* of the signal s. To analyze the frequency content of the signals in the context of the graph frequencies, we utilize the concept of *total variation on graphs* from DSP_G (Sandryhaila & Moura 2014b), which provides a mathematical basis for ordering frequencies. In classical discrete signal processing, the total variation of a discrete signal is defined as the sum of magnitudes of differences between consecutive signal samples. Total variation applied to arbitrary graphs, such as the graph defining occupant relationships, is determined by the eigenvalues of the adjacency matrix $\mathbf{A}$. The total variation of an eigenvector $\mathbf{v}_n$ of a matrix $\mathbf{A}$ is:

$$TV_G(\mathbf{v}_n) = \left|1 - \frac{\lambda_n}{|\lambda_{\max}|}\right| \|\mathbf{v}_n\|_1 \tag{4.9}$$

where $\|\mathbf{v}_n\|_1$ = L1-norm of the eigenvector $\mathbf{v}_n$. The TV_G value for each normalized proper eigenvector is between 0 and 2. Theoretical analysis of the TV_G concept can be found in (Sandryhaila & Moura 2014b).

By sorting frequencies from low to high by their total variation, the variability associated with the differences in weighting between nodes becomes accessible. *If a signal's frequency content is concentrated in the lower frequencies, the variation in the signal's values follows the weighting pattern of the graph (i.e., two nodes with a relatively high weighting between them would have relatively similar expected signal values).* When signals from sensors across a spatial and social domain are expected to have little variability (as would be the case when occupants who are both near each other and part of the same organization are using relatively similar amounts of energy), the graph spectral plot would be expected to have this characteristic shape. However, with more variability across nodes with large edge weightings, the signal would have more of its energy in the higher frequencies. This change in the graph spectral plot could allow for potential flagging of unexpected occupant energy use in a given building or floor plan.

Figure 4.1 depicts the overall flow of the $\mathrm{OESP_G}$ framework. First, physical locations of occupant workstations as well as the organization or team to which the occupant belongs to are recorded (1). We note that the framework can be utilized for lower spatial resolutions (e.g., groups of desks), but a key strength of the $\mathrm{OESP_G}$ framework is its ability to efficiently process high spatial resolution data. This spatial and social information is used to construct a graph that describes the underlying structure of the building (2). The computationally intensive part of the framework is the eigendecomposition of the graph's adjacency matrix (3), which allows the structure of the building's occupant network to be decomposed into characteristic frequencies and characteristic frequency components that describe variability across the graph structure—with higher frequencies indicating localized areas of higher signal variability across the constructed graph. The eigendecomposition results in the graph Fourier transform matrix (4), and using the concept of total variation on graphs (5), the frequency spectrum can be ordered from high to low (6). The eigendecomposition of the adjacency matrix need only be done once, allowing the framework to easily scale to large buildings—and even districts of buildings—with thousands of occupants. Once the graph describing the spatial and organizational layout of occupants in the building has been defined and decomposed, energy use data collected through plug load sensors can be analyzed in the spectral domain. The plug load sensors collect snapshots of power usage at regular intervals, which become the signals in the $\mathrm{OESP_G}$ framework (7). The iterative aspects of the framework involve normalizing the data (8-9) and multiplying the normalized signal (10) with the graph Fourier transform matrix to determine the frequency content of the signal (11). This process allows for the creation of the frequency plot (12), which can be analyzed to understand spatial, temporal, and social dynamics of each energy-use signal. As long as data is being collected (13), new signals can be defined, and new frequency plots can be created at each period.

4.4 Simulated example

In this subsection, we present a simple simulated example to elucidate the core concepts of $\mathrm{OESP_G}$ and demonstrate its applicability to identifying potential anomalies in energy use across the floorplan of a building.

4.4.1 Data simulation

The floorplan, social network structure, and energy use data are all simulated in this example. The relative locations of the simulated plug load sensors are shown in Figure 4.2. Eight sensors are used,

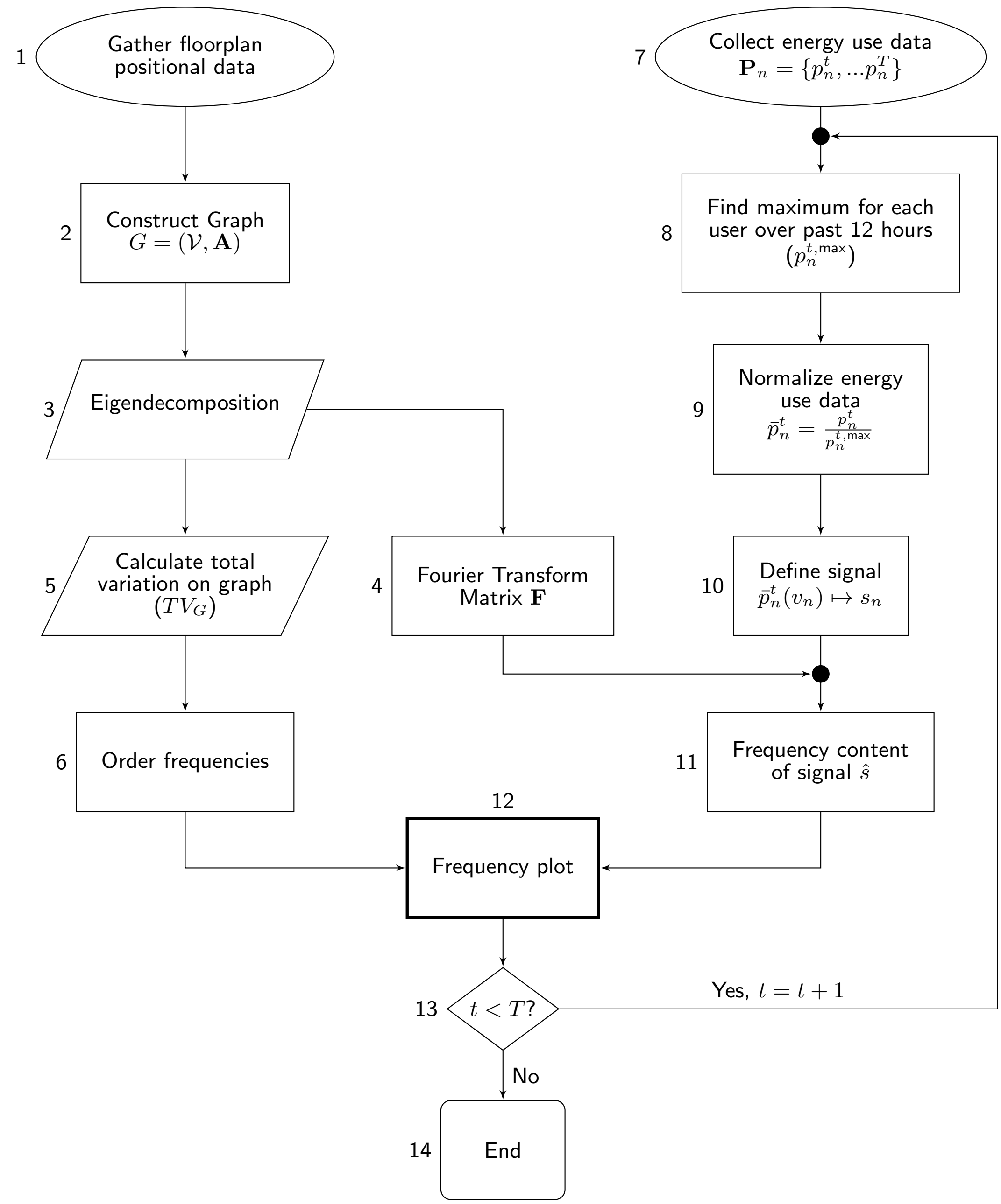

Figure 4.1: $\mathrm{OESP_G}$ framework flow.

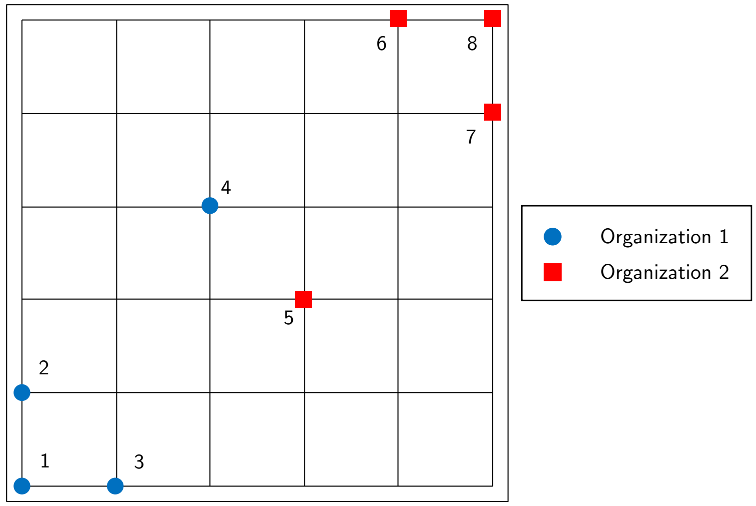

Figure 4.2: Simulated floor plan with plug load sensors in three spatial clusters and two organizations (circles and squares).

with three clusters of individuals. The two corner clusters each contain three individuals sitting near each other, with each of the two clusters belonging to a different organization. The cluster of circles on the bottom left of the figure is organization 1 (blue circles), and the cluster of squares on the top right is organization 2 (red squares). The third cluster of two individuals, located directly in between the two corner clusters, contains one individual from each organization.

Energy use data streams for each sensor provide a snapshot of power usage for the sensor at 20-minute intervals. The simulated sensors capture data for occupant workstations that can be in one of three states: working, on break, or not present. Values for these states are derived from a study by Lawrence Berkeley National Lab on representative power draw from commonly used office desk equipment. The purpose of assigning typical real power draw values to occupant states is to simplify the simulated example for illustrative purposes. The "not present" state is assigned to 2W of power draw (roughly corresponding to a laptop and monitor in off mode), the "on break" state is assigned to 10W of power draw (roughly corresponding to a computer display and laptop in sleep mode), and the "present state" is assigned to 50W of power draw (roughly corresponding to a laptop and monitor in awake mode) (Lawrence Berkeley National Laboratory 2016). The standard work schedule is chosen as 9 a.m.-5 p.m., with an hour lunch break at 12 p.m. Variations on this schedule are used to test how spatial, temporal, and social variations in occupant energy use can be captured using the proposed framework.

4.4.2 Graph construction

Using the simulated locations of sensors on a floorplan as the basis for a graph, the adjacency matrix can be calculated. The graph is constructed as an undirected weighted graph wherein each node is connected to all other nodes. Edge weightings are calculated through two components: (1) the Gaussian weighting function capturing the physical distances between sensor locations, and (2) a binary function capturing the organizational identity of the sensor and its associated occupant. For two nodes n and m, the graph weighting is found as:

$$\mathbf{A}_{n,m} = \mathbf{A}_{m,n} = e^{-\frac{d_{n,m}^2}{2\sigma^2}} + \alpha f_s(n,m) \tag{4.10}$$

where $d_{n,m}$ = Euclidean distance between the nodes, the Gaussian standard deviation σ is a user defined parameter that controls the width of the distribution (for the purpose of this example, we assume the standard $\sigma = 1$), and the function $f_s(n,m)$ describes the social network relationship between the two nodes, with $f_s(n,m)$ taking a value of 1 if the two occupants are in the same organization, and 0 if the two occupants are not in the same organization. In this example, therefore, the social structure is modeled directly after each occupant's organizational identity, but we note that other relationships, including those at the sub-organization level, can be utilized to build the social component of the edge weightings. Additionally, the social dimensions could include other relationships beyond institutional identity, such as networks of friends or social groups. However, for this analysis, the social dimension is limited to organizational identities. The parameter α weights the importance of the social dimension of the graph structure, with a higher value indicating that the social dynamics are expected to be of higher importance. For our example, we assign $\alpha = 1$, indicating that social dynamics are of the same importance as spatial dynamics.

The spatial component of the graph construction gives a larger weighting to edges connecting nodes that represent sensors physically near each other, and smaller weighting to edges connecting nodes that represent sensors far away from each other. The intuition behind this notion of edge weighting comes from the expectation that individuals sitting near each other are more likely to have similar energy use and occupancy patterns compared with individuals sitting far apart from each other. Additionally, the social component of the graph construction gives larger weighting to edges connecting nodes that are part of the same organization, following the intuition that people of the same network would be expected to have relatively similar energy use and occupancy patterns.

4.4.3 Analysis

Given the eigendecomposition of the adjacency matrix, as found in eq. 4.7, each energy use signal can be expanded into the graph spectral domain, following the OESP_G framework process described above. Figure 4.3 shows the graph spectral plot for each of four possible schedules in the simulated building. In the schedules shown in the figure, dark green is associated with the *working* state, light blue with the *on break* state, and white with the *not present* state. The graph spectral plots to the right of the schedules show the frequency content for the 12 p.m. signal. This signal is chosen because the simulated schedule shifts have impacts on which occupants are on break and which are at their desk at 12pm. Figure 4.3a is the baseline scenario in which each occupant in the simulated building has the same schedule: start work at 9 a.m., take a break from 12 p.m. to 1 p.m., and leave at 5 p.m. Figure 4.3b-d represents scenarios in which shifts by one or more individuals are made according to the associated schedule. The spectral analysis for each scenario indicates that the change in schedule has impacts on signal frequency content. A sensitivity analysis on the parameters introduced in eq. 4.10 shows that varying either σ or α has little effect on the frequency plots. In the sensitivity analysis, we allowed σ and α to change by multiplying or dividing by 2, and in all cases, we observed the same overall trends as shown in Figure 4.3, in which $\sigma = 1$ and $\alpha = 1$. After the running the sensitivity analysis on scenario (d), the maximum change for the lowest frequency was 0.7% and the maximum change for two highly expressed high frequencies (indexed 7 and 8) was 12%.

4.4.4 Simulated example results and discussion

When all eight simulated plug load sensors follow the same schedule, the graph spectral plot indicates that the analyzed signal's frequency content is concentrated in the lowest frequencies, as shown in Figure 4.3a. When one individual shifts his or her schedule, as tested in Figure 4.3b, the graph spectral plot shows an increase in signal energy in the higher frequencies. This increase in high frequency energy is caused by the now-incongruous energy use patterns between the shifted individual (occupant 1) and the two non-shifted individuals in the same cluster who are closely related to occupant 1 both spatially and socially (occupants 2 and 3).

Figures 4.3c and 4.3d show two examples in which all of organization 2 shifts along with one member from organization 1. Occupants 5-8, who comprise all of organization 2, all shift their schedule by one hour in both scenarios. When occupant 4 shifts with them, the graph frequency plot shows increased power in the middle frequencies. When occupant 1 shifts with them, the graph frequency plot shows increased power in the higher frequencies. This result makes sense given that occupant 4, while engaging in behavior different from the rest of his or her organization, is both more spatially related to organization 2 and less spatially related to organization 1 than is occupant 1. When occupant 1 shifts, the result is similar to that from the situation depicted in Figure 4.3b. If we are interested in detecting situations that could lead to recommendations for more efficient building management, this analysis can provide insight into subtleties associated with complex occupant behavior. The situations in (c) and (d) seem very similar, yet it becomes clear from this analysis that the spatial incongruity in (d) would make it impossible to implement energy-saving strategies such as reduced HVAC service to a zone encapsulating the cluster of occupants 1, 2, and 3.

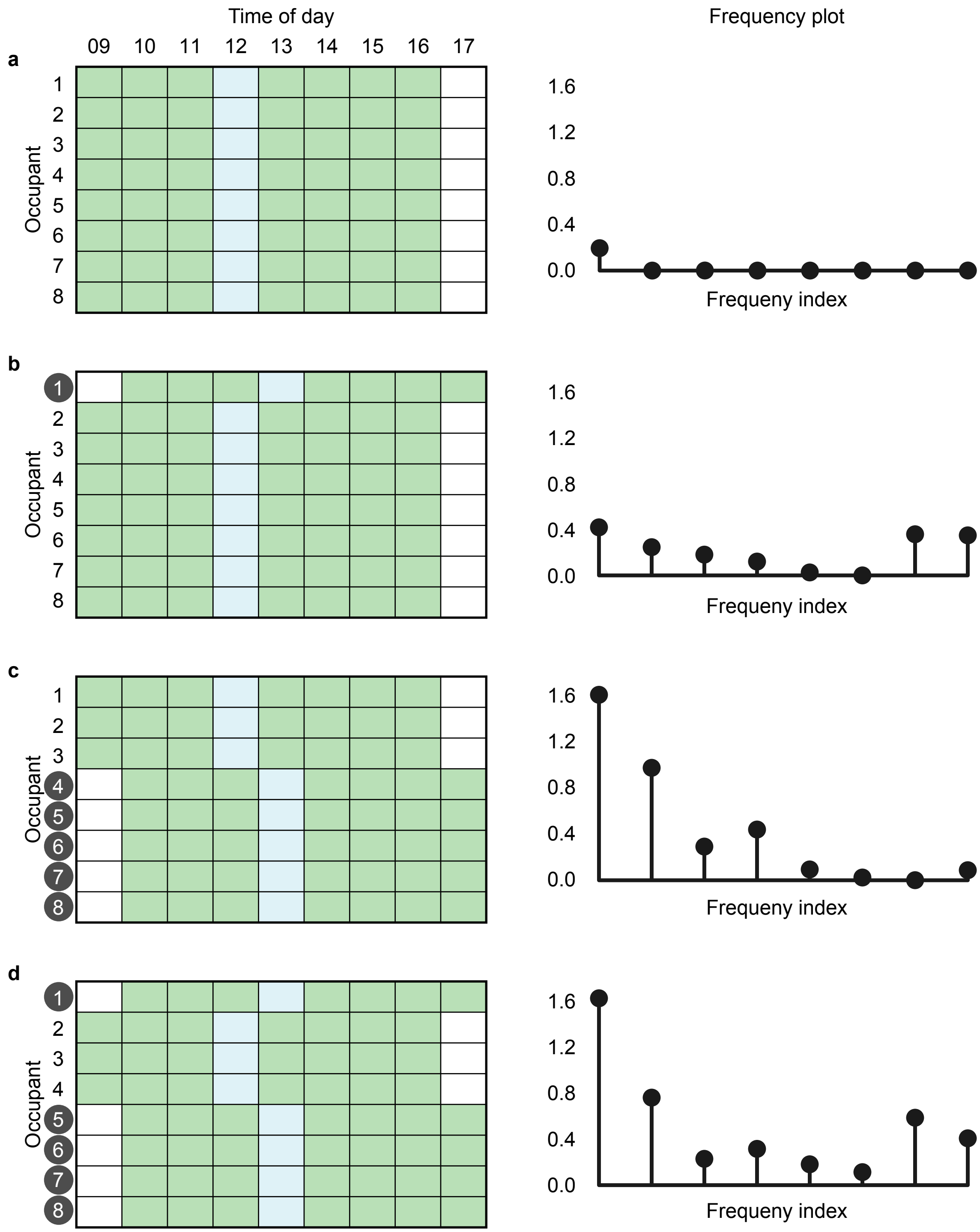

Figure 4.3: Schedules and graph spectral plots for simulated example.

A single detection of this incongruity could lead to recommendations for schedule shifts that more closely align spatially-related individuals of the same organization or across organizations, allowing for potential energy savings. Repeated detections could also lead to recommendations for spatial adjustment of occupants, which could also lead to potential energy savings. This simulated example illustrates the ability of the OESP_G framework to identify incongruities along the three dimensions of occupant energy use that it analyzes, by detecting large values for the higher frequencies in the frequency plot. The example highlights the power of the framework in terms of detecting situations in which recommendations for energy efficiency strategies could make a real impact on a building's performance.

4.5 Case study: office building in Denver, Colorado

In this section, we apply the OESP_G framework to analyze real data and formulate strategic recommendations for energy efficient operations of a case study office building in Denver, Colorado.

4.5.1 Data collection and normalization

Data was collected using off-the-shelf plug load monitors (i.e., Monster Cable 300MC PowerControl unit) installed at individual desks throughout two floors of an existing and occupied 6-story office [$3716m^2$ ($40,000ft^2$)] test-bed building in Denver, CO. In this building, employees were typically present between 9:00 a.m. and 5:00 p.m. from Monday to Friday. Workstations most often included computers, monitors, space heaters, and electronics chargers, and these appliances were connected to the plug load monitor through a power strip. The Monster Cable 300MC plug load monitors connected to standard North American 120 V outlets and communicated information to the included Monster Cable edge-router (GTW 100) that uploaded data to a database via an Ethernet based internet connection. Real time power draw (W) was collected at 20-minute intervals. More information regarding the Monster Cable 300MC plug load monitoring equipment specifications and test-bed building set-up can be found in Gulbinas and Taylor (2014) and on manufacturer's retail website (http://www.monstorproducts.com). Within the two floors of the building, data was collected for a total of 27 individuals' workstations in 5 separate organizations. The physical location of each sensor as well as the organizational association of each individual was recorded to indicate the spatial and social attributes associated with each workstation; these attributes are shown in Figure 4.4. The color of the sensor on the test-bed building floorplan refers to the organizational identity of the occupant associated with the sensor, with each color representing one organization.

We captured power use at 20-min intervals for each workstation in the study. Typical values for power use ranged from 50W to 200W for the workstation, depending on the appliances plugged into the power strip. To account for this range in absolute values of power draw, we normalized each workstation's power use over the previous 12 h of data collection, as described in the Methodology

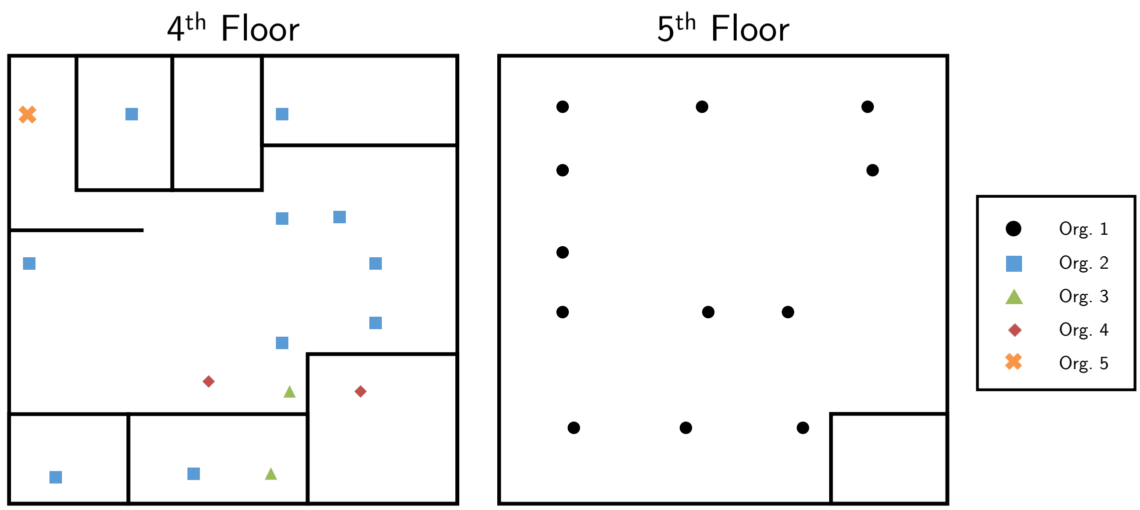

Figure 4.4: Workstation locations superimposed on building floorplans.

section. This method allowed for comparisons among individuals' relative energy use behavior over the course of a single work day.

4.5.2 Case study: results and discussion

By applying the $\mathrm{OESP_G}$ framework to this data, we were able to analyze the variability in energy use behavior in terms of spatial layout and social network structure across the floor plans of the building. Both physical distances and organizational affiliations were used to construct the adjacency matrix, following eq. 4.10 above. Spikes in frequency content for high frequencies are of interest because they indicate instances of high variability (i.e., points in time in which individuals are not drawing power as one would expect). These expectations are embedded in the graph construction. We would expect occupants with similar spatial characteristics (i.e., those sitting close to one another) to have similar energy use patterns, and similarly, we would expect occupants with similar social characteristics (i.e., those that are part of the same organization) to have similar energy use patterns. In general, occupants with similar characteristics have higher edge weightings between them. When similar occupants have distinctly different energy use patterns, their energy use behavior can be considered *out-of-sync* with expectations.

Using this framework, we can apply a threshold to the higher frequencies. When the frequency contents of the higher frequencies cross the threshold, occupant energy use behavior is deemed *out-of-sync*. For the purposes of this case study, we utilize simple heuristics from previous work (Sandryhaila & Moura 2014a) to indicate the high frequencies of interest to be the half at the top of the spectrum (14-27 in this application) and the frequency content threshold to be 1 on the y-axis. Using this simple threshold, certain *out-of-sync* signals can be detected and analyzed. Figures 4.5a

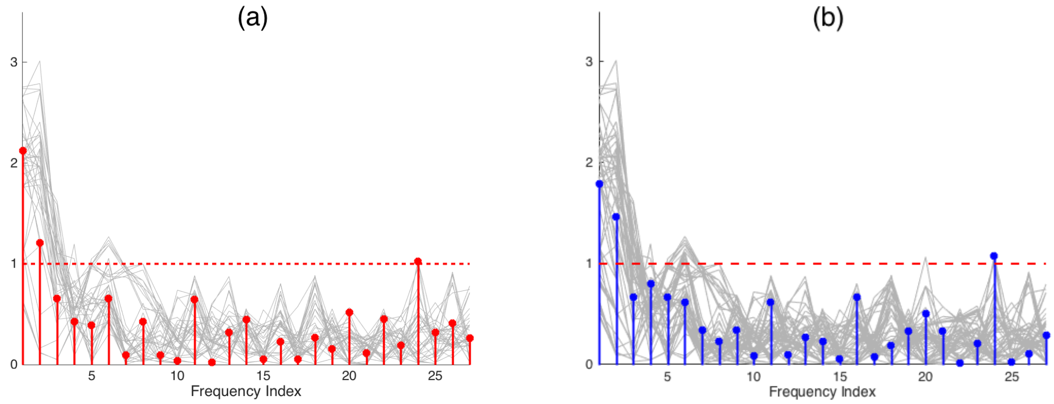

Figure 4.5: Frequency plots for one representative workday, with detected signals: (a) 2:40 p.m.; (b) 3:00 p.m.

and 4.5b show the frequency plot of each signal over the course of a full workday in gray, as well as one signal that is detected as *out-of-sync*. In Figure 4.5a, the detected signal, in red, occurs at 2:40 p.m., and in Figure 4.5b, the detected signal, in blue, occurs at 3:00 p.m.

In this particular example, two signals are detected, one right after the other. The first occurs at 2:40 p.m. on the analyzed workday, and the second occurs at 3:00 p.m. on the analyzed workday. Since the signals are captured at 20-minute intervals, this detection could indicate that 40 minutes of the workday are in the *out-of-sync* condition. The high frequency that caused these detections is the 24th, as can be seen in Figures 4.5a and 4.5b. To understand the cause of the out-of-sync condition as detected in this analysis, the eigenvector associated with the 24th eigenvalue—as ordered by the total variation—can be plotted and its components can be analyzed (Figure 4.6). Analyzing this eigenvector provides insight into which nodes are responsible for the signal detection (Deri & Moura 2015). The figure shows that nodes 5, 6, 7, and 8—nodes that are both close to one another and part of the same organization—are most highly expressed in this 24th frequency. With a relatively high amount of power in this high frequency, it would be expected that the highly expressed nodes in the corresponding eigenvector would exhibit incongruous power draw behavior. In this example, the power values at nodes 5-8 describe a situation in which power values for nodes 5 and 8 rapidly become small compared to recent patterns, while power values for nodes 6 and 7 are near the maximum amount of power drawn recently. Specifically, at both detected signals, occupants 6 and 7 are both drawing more than 80% of their individual maximums (as iteratively measured over the previous 12 hours), while occupants 5 and 8 are both drawing 0% of their individual maximums (as iteratively measured over the previous 12 hours). This data is summarized in more detail in Table 4.1.

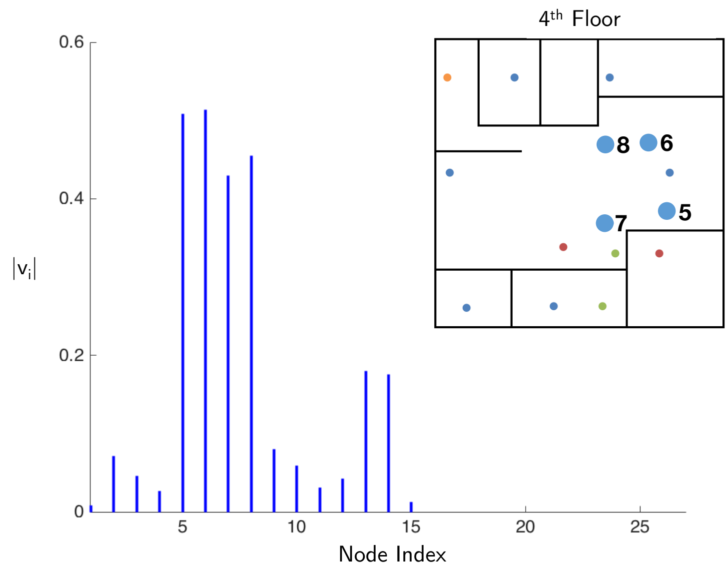

Figure 4.6: Components ($|v_i|$) of the eigenvector of the 24th frequency of the plot in Figure 4.5 (detected as *out-of-sync*).

Table 4.1: Summary of power draw values for *out-of-sync* signals.

Occupant	Normalized power draw, 2:40 p.m. (fraction of max)	Normalized power draw, 3:00 p.m. (fraction of max)	Mean normalized power draw over full day (fraction of max)
5	0	0	0.22
6	0.87	1.00	0.39
7	0.87	0.87	0.40
8	0	0	0.25

4.5.3 Energy efficiency recommendation strategy

Using power values as a proxy for occupant behavior, we can draw the conclusion that the two groups of two individuals are following different schedules, resulting in a situation in which building energy use for things like lighting and thermal comfort might not be as efficient as possible. If a recommendation can be made such that all four occupants follow the same schedule for the day, the signals that were once detected are no longer detected (Figure 4.7). The figure shows much lower energy exhibited by the 24th frequency for the same signals that caused the energy in the 24th frequency to exceed the learned threshold we had set. By making a schedule-shifting recommendation for the occupants so that all occupants within a zone of the building follow the same schedule over the course of a day, we are able to show that the high frequencies in our framework are sensitive to recommendations that more closely align individuals who are expected to engage in similar behaviors.

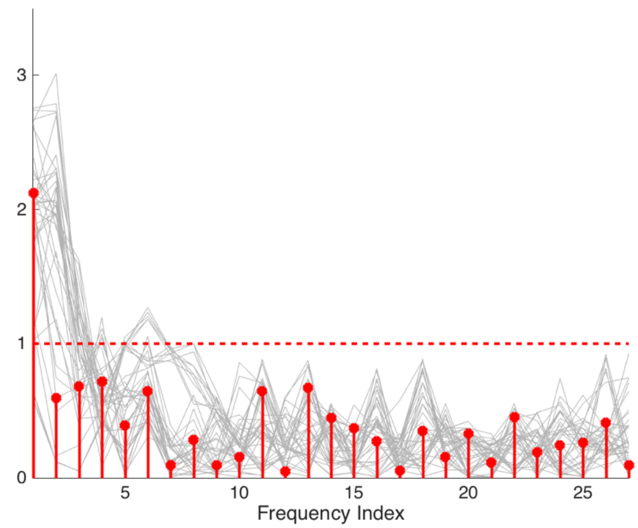

Figure 4.7: Frequency plot for one workday after schedule recommendations.

In this particular example, the power of the framework is in recognizing points in a workday in which nuanced behavioral dynamics of closely related individuals (by space and social network) are less aligned than would be expected. When situations like these are detected, there exist opportunities to align schedules and take advantage of granular building system controls. Distributed and highly controllable building systems will operate most effectively when complex occupant behavioral dynamics are best understood. This framework introduces a methodology for understanding how occupants use a building and how that behavior differs from our expectations, in terms of spatial and organizational correlations. Using this understanding to improve distributed and precise building controls could lead to large potential energy savings. One important advantage of the $\mathrm{OESP_G}$ framework is its scalability. Very large buildings with many groups of individuals, many floors, and many occupants can be analyzed quickly after the initial work of defining adjacencies and determining frequencies through the Fourier analysis. Signals with relatively high energies in the high frequencies—and the nodes responsible for this result—can be identified in real time for quick recommendations.

4.6 Limitations and future work

The main limitations for this initial study of the novel $\mathrm{OESP_G}$ framework include parameter fitting and recommendation strategies. When assigning the edge weightings that comprise the adjacency matrix, we used typical values to simplify the analysis. Future work should consider a methodology for finding the best weighting scheme for both the Gaussian distance weighting and the organizational weighting. We also note the potential limitation of combining the spatial and organizational components of the building's organizational structure into one weighting. While separating these factors would mean sacrificing the ability of the framework to analyze all three dimensions simultaneously, a focused spatial or social analysis might provide additional insight. Additionally, there is potential for future work to investigate the threshold that is applied to the higher frequencies in order to determine the *out-of-sync* condition. While our method is consistent with heuristics from previous work in $\mathrm{DSP_G}$, further understanding of appropriate thresholding for the domain specific area of occupant analysis could further improve the efficacy of occupant analysis frameworks like $\mathrm{OESP_G}$.

One other area of future work could involve improving the signal input vector by constructing a composite signal based on numerous other data sources in addition to plug load monitors. Other data collection devices, such as occupancy sensors, could add information beyond what is available from collecting only energy consumption data. While these data streams could provide additional information valuable in the analysis of occupant dynamics, they may introduce additional uncertainties and challenges in regards to scalability, reliability and fusion of disparate data streams. Therefore, we leave the development of such a composite signal for future work.

There is also potential for future work that considers how best to make recommendations for occupant schedule and spatial shifts in order to both reduce high frequency energy and to ultimately improve building energy performance. This work could include algorithms for building systems, building management practices, and occupant feedback tools. Robust recommendation strategies would create a link between the identification of potentially problematic occupant behavior (what our framework accomplishes) and better building energy performance. Future studies might also consider a scope beyond that of energy use in a single building envelope. As more districts and cities begin collecting live energy use for buildings, the inherent scalability of the $\mathrm{OESP_G}$ framework allows for a much larger scale of analysis. An additional flexibility of the framework is that its signal need not be limited to power or energy. Future work might look at other sustainability indicators, such as pedestrian or automobile traffic flows.

One exciting potential area of research that builds off this framework is the inverse problem considered in this paper. That is, given the spatial layout of occupants and a dataset describing their energy use behavior, could the inherent social structure of the building be inferred by minimizing the energy in the high frequencies of the signals' frequency plots over time? Such an analysis would provide valuable insight into how social networks form within a building given organizational identity, spatial configuration, and energy use. This insight could enable the design of new buildings that maximize occupant interaction and minimize energy usage.

4.7 Conclusions

The primary purpose of this paper was to introduce and test the OESP_G framework, a data framework grounded in the emerging area of signal processing on graphs and capable of analyzing occupant behavior in three core dimensions: *spatial*, *temporal*, and *social*. Extending previous data-driven occupant-based analysis frameworks (Azar & Menassa 2016; Gulbinas et al. 2015), a major contribution of the OESP_G framework is its ability to simultaneously analyze data across the three key dimensions of occupant energy use within a commercial building. By using the physical locations and organizational or team identity of individuals and their workstations, we define a graph with edges between nodes that are weighted based on these spatial and social dimensions. Using power draw signals from the occupant workstations, the OESP_G framework analyzes the variability of the signal across the constructed graph, identifying in real time situations in which occupants behave differently from other occupants closely related by space and social structure. These incongruities are detected as spikes in the high frequencies of the frequency plot, which indicate high variability across one or more dimensions. The OESP_G framework can be utilized to provide insight into which occupants are responsible for this high variability across the graph, and, using this information, it can yield simple recommendations to more closely align individuals and enable more energy-efficient operations of building systems. These strengths are demonstrated in both a simulated and real case study example.

In addition to addressing the multi-dimensionality problem associated with commercial building energy data, the OESP_G framework was designed to be scalable to very large buildings with thousands of occupants. The underlying graph structure and computational efficiency of the single eigendecomposition lends itself to the efficient real-time analysis of large commercial buildings with thousands of occupants and even multiple buildings. As a result, the proposed OESP_G framework is a building block for more efficient data-driven management of building systems, better recommendations for occupant behavior, and even better design of building layouts for improved energy efficiency.

By utilizing new energy use data streams, a deeper understanding of the complexity of interactions among the various dimensions of occupant energy use in buildings has the potential to yield significant energy savings in commercial buildings and enhance occupant comfort of spaces. Given the large role of buildings in the energy use landscape, data-driven efficiency strategies for commercial buildings will prove to be invaluable in addressing modern day environmental crises and meeting our sustainability goals.

Chapter 5

Data-driven optimization of building layouts for energy efficiency

This chapter is adapted from the following paper: Sonta, A., Dougherty, T. R., and Jain, R. K. (*submitted*). "Data-driven optimization of building layouts for energy efficiency."

Abstract

One of the primary driving factors in building energy performance is occupant behavioral dynamics. As a result, the layout of building occupant workstations is likely to influence energy consumption. In this paper, we introduce methods for relating lighting zone energy to zone-level occupant dynamics, simulating energy consumption of a lighting system based on this relationship, and optimizing the layout of buildings through the use of both a clustering-based approach and a genetic algorithm in order to reduce energy consumption. We find in a case study that nonhomogeneous behavior (i.e., high diversity) among occupant schedules positively correlates with the energy consumption of a highly controllable lighting system. We additionally find through data-driven simulation that the nave clustering-based optimization and the genetic algorithm (which makes use of the energy simulation engine) produce layouts that reduce energy consumption by roughly 5% compared to the existing layout of a real office space comprised of 165 occupants. Overall, this study demonstrates the merits of utilizing low-cost dynamic design of existing building layouts as a means to reduce energy usage. Our work provides an additional path to reach our sustainable energy goals in the built environment through new non-capital-intensive interventions.

5.1 Introduction

The energy performance of buildings is largely driven by the operation of their energy-intensive systems. In commercial office buildings, the most energy-intensive systems are those that provide comfortable thermal and visual environments (i.e., heating, cooling, and lighting systems). The operation of these systems critically depends on the subjective experience of building occupants when they are using a building's spaces. As a result, there is no need to heat, cool, or light spaces that are not used at a particular point in time. These thermal and lighting systems are often controlled by zone, and in the case that even one occupant enters a zone, the systems must typically service the entire zone. This shared feature of building system operation contributes heavily to inefficiency (Yang et al. 2016), but it also creates the opportunity to optimize the design and management of building spaces and save energy through the individualization of building spaces.

Let us consider a hypothetical example of an office building with 4 teams, 4 members per team, and 4 shared rooms/zones. In this office, for the purposes of this example, one member of each team must be present at all times. All 4 teams have decided to operate on the following schedule: the first team member works from 12 a.m. – 6 a.m., the second from 6 a.m. – 12 p.m., the third from 12 p.m. – 6 p.m., and the fourth from 6 p.m. – 12 a.m. . If the office is arranged such that each team occupies its own room/zone, there will be one person in each room at all times. In other words, all 4 rooms will have exactly 1 occupant in the room at all times. We refer to this situation as an example of high *occupant diversity* within each building zone, where diversity is a term used to describe the differences in activities or schedules. As a result of these differences, the heating, cooling, and lighting systems for all four rooms will operate at all times. We could instead arrange the layout such that each occupant shares a room with their fellow shift-workers (e.g., all 12 a.m. – 6 a.m. workers share a room). In this case, only one room, the occupied room, will need to be supplied with heating, cooling, and lighting throughout the day. The operation time of these systems would therefore be reduced by 75% compared to the first scenario.

While this is an extreme example, the underlying dynamics apply to all buildings with shared spaces. In reality, the complexities and subtleties of occupant schedules, particularly in large office buildings, make it difficult to discern shared patterns of behavior. The decisions of creating building layouts, therefore, typically involve functional or hierarchical structures of organizations. In this paper, we investigate the possibility of optimizing layouts based on occupants' use of spaces.

Past research into building energy efficiency has focused on both building design as well as building operation. A large body of work has focused recently on the impact of occupant behavior on energy-intensive operation of building systems. Furthermore, researchers have investigated the role of design optimization, including through an occupant-centric lens, in reducing building energy consumption. Below, we discuss key findings from previous research and motivate the work presented in this paper.

5.1.1 Building design and energy efficiency

It is well known that the design of buildings has a large impact on their future operation, including energy efficiency (Lucon et al. 2014). When considering energy in building design, architects and engineers generally consider physical building parameters including orientation, materiality, fenestration, and choice of heating, cooling, and lighting systems (Basbagill et al. 2014; Petersen & Svendsen 2010; Roisin et al. 2008). The design process also considers, in addition to these physical characteristics of individual components, the impacts of layout on occupant behavior through the architectural programming process. This programming often is driven by intended building use (e.g., meeting rooms or workspaces in an office building). While this design process is typically regarded in terms of our subjective experience of the building, past research has shown that these choices for building layouts can influence the energy consumption of the building (Yang et al. 2016). For example, locating parts of a home that are more often used in the morning on the eastern side of the house could reduce the amount of heating, cooling, and lighting required for those spaces.

An advantage of considering the layout of the building in the context of energy efficiency is that layouts are often flexible, especially compared to other design considerations such as orientation and materiality. This is especially important due to the fact that our use of buildings evolves over time. Buildings originally intended for one purpose are often repurposed to suit changing needs, leading to different patterns of space utilization and rendering some physical design considerations obsolete. Moreover, as buildings lifespans are on the order of 40-100 years (Tanikawa & Hashimoto 2009), it is expected that the majority of energy consumption from the building sector will come from existing buildings rather than new construction for many upcoming years—years that are critical to our sustainable energy goals (Lucon et al. 2014). This importance of the existing building stock in addressing energy challenges suggests the need for new and innovative ways to reconsider the energy-intensive attributes of existing building design. New design methods that focus on building layouts are a promising means of addressing these challenges.

5.1.2 The role of the occupant in energy efficiency

As discussed above, building layouts are often driven by intended use. Research has shown that the actual use of spaces by building occupants, while extremely impactful to building energy consumption, is both very difficult to model and understand (Feng et al. 2015; Parys et al. 2009; Roetzel et al. 2014). For example, a seminal study showed that changes in occupant behavior can cause two-to-one discrepancies in actual versus expected energy consumption (Norford et al. 1994). Both occupant actions (e.g., interactions with building systems through thermostats, windows, etc.) as well as their passive use of space are important to energy consumption (Jia et al. 2017). This latter notion, passive space use, largely drives energy consumption in buildings that have systems that are able to respond (e.g., turn off or reduce service) based on occupancy information. These responsive systems are increasing in prevalence as energy codes and regulations are driving wider adoption of energy-efficient technologies in the building sector (Park et al. 2019).

5.1.3 Optimizing buildings for energy efficiency

As both design and occupant behavior are key factors that affect the energy efficiency of buildings, researchers have sought to leverage optimization of building design and operation as a tool to improve energy performance. These optimization approaches have generally focused either on building design pre-construction or operation of heating, cooling, and lighting systems post-construction. Design optimization is typically considered in the context of new construction (Basbagill et al. 2014). Such research on design optimization has largely focused on physical building parameters in the early stages of design (Basbagill et al. 2014; Yu et al. 2015). However, because of the *coupled* effects of occupant behavior and building design on building performance, researchers have noted that occupant-related uncertainty can hinder confidence in early-stage design decisions supported by such optimizations (Hoes et al. 2011). These coupled dynamics have driven researchers to focus more on the operational phase of building—and particularly the importance of building occupants as a means of improving energy efficiency (Park et al. 2019).

Researchers have therefore developed data-driven tools that optimize control of energy-intensive systems in existing buildings, finding that optimizing control of systems through an occupant lens can enable large reductions in energy consumption, ranging from 15-70% depending on the types of systems analyzed (Agarwal et al. 2010; Balaji et al. 2013; Krioukov et al. 2011). Recently, researchers have found that an important component of such control optimization strategies is the explicit consideration of occupant comfort, which improves the subjective experience of the occupants but can reduce the energy savings possible (Ghahramani et al. 2018). This research shows the promise of introducing sustainable energy savings in existing buildings by controlling building systems optimally. However, this research also considers the dynamics of occupant behavior as given. To take full advantage of controllable building systems, there remains the opportunity not only to optimize how systems respond to building space use, but also to optimize how people themselves use spaces—an aspect we consider in this paper through the building layout.

The previous research discussed in this section has shown that we can optimize building designs before construction and optimize building controls after construction. However, the dynamic behavior of building occupants—behavior that enables research into building controls—can also enable new research into building design features that remain flexible once the building is erected. As discussed above, a focus on building layouts may offer a means for addressing energy gaps in design through a naturally occupant-focused lens. A key research gap, therefore, is the integration of design optimization with important characteristics of dynamic occupant behavior. While optimization of physical building parameters before construction will remain a promising area of inquiry, there is a need to be able to dynamically examine aspects of building design once its key operational parameter, the building occupant, enters the picture. Optimization of building layouts have been considered in terms of organizational structure and performance (Jo & Gero 1998; Lather et al. 2019; Lee et al. 2012), however, the direct optimization of building layouts to address the energy-efficient operation of building systems remains an area for continual research.

In this paper, our overarching research question is whether building layouts can be optimized to reduce energy consumption of energy-intensive systems by leveraging sensor data on occupant activities. We first describe our previously introduced methods for abstracting time series plug load data to states of occupant behavior. We then discuss our methods of analyzing zone-level diversity in occupant schedules, optimizing layouts, and simulating the impacts of adopting an optimized layout. We introduce a real-world case study, where we apply our methods to a floor of an office building with 165 occupants.

5.2 Methodology and data collection

In this section, we describe our overall methodology for developing a framework that enables the optimization of building layouts for energy savings (Figure 5.1). We first leverage ambient sensor data collected from plug load energy sensors at the individual desk level to describe occupants' use of space (see section 2.3 above). We term this description the individual's *occupant schedule*. We then define a distance metric that can be used to describe the *zone diversity* in occupant schedules over several individuals (section 5.2.1). This distance metric creates the ability to cluster occupants spatially (i.e., create new layouts) in an effort to reduce this zone diversity and therefore building energy consumption (section 5.2.2). Due to the extremely high dimensionality of the possible solution space, we also introduce a genetic algorithm for creating occupant layouts based on expected energy consumption (section 5.2.3). For evaluation of our layout optimization algorithms, we introduce a data-driven surrogate model for simulating energy consumption of a building's lighting system based on occupant schedules (section 5.2.4). We describe the dataset used for analysis in section 5.2.5.

5.2.1 Representing zone diversity of occupant schedules using Euclidean distance

AAs discussed above, buildings provide energy-intensive services by zones, and the spatial efficiency of providing these services depends on occupant schedules. Therefore, a key question in understanding the operation of these systems, from the perspective of spatial efficiency, concerns the similarities or differences among the schedules of occupants within individual zones. We term these similarities and differences as the *diversity* in occupant schedules among occupants within a given space, and we operationalize this measure on time series data. Based on the work in Yang et al. (2016), we can define this diversity as the distance between the vectors describing time series schedules for each

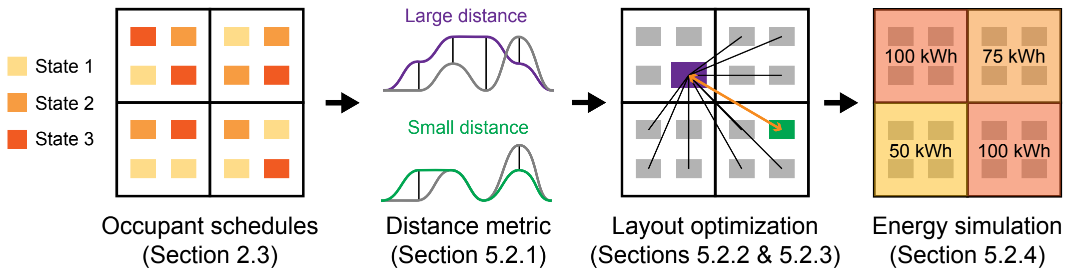

Figure 5.1: Outline of methodology.

occupant in the zone. A range of distance metrics could be used, including cosine similarity, Manhattan distance, Euclidean distance, etc. Following previous research practice (Yang et al. 2016), we use the Euclidean distance for this study, but we note that the specific distance choice does not have a large impact on the analysis.

If our schedule data is structured as above: $\mathbf{S}_{i,t}$, where i is the occupant index and t is the time index, we can compute distances between the schedules for any two occupants. For example, the distance between occupant i and occupant j can be computed as follows:

$$d_{i,j} = \sqrt{\sum_{t=0}^{T} (\mathbf{S}_{i,t} - \mathbf{S}_{j,t})^2} \tag{5.1}$$

Using this distance metric, we can compute the distances between all occupants in a zone, forming a distance matrix. Normalizing this distance matrix by the total number of entries in the matrix (except the diagonal, since occupants' distance from themselves is 0), we have an average distance among all the occupant schedules within the zone, which we define as the *overall zone diversity*. With this metric, we can compare the diversity of occupant schedules for individual building zones to the actual energy consumption of the building systems. We would expect higher zone diversity to correlate with higher energy consumption, as was shown using physics-based simulation in Yang et al. (2016).

5.2.2 Optimizing layouts: Dimensionality reduction and occupant clustering

Dimensionality reduction using Singular Value Decomposition

Given occupant activity states and the notion of zone diversity in occupant dynamics, our next objective is to create optimal groupings of occupants in space—that is, to optimizing a building's layout. Time series sensing generally produces many signals over time for each sensor deployed. In our case, activity states are generally reported on the scale of 15 minutes, creating 96 signals per occupant per day, or up to 35,000 signals per year. Computation of distances between vectors of this size suffers from the well-documents curse of dimensionality, whereby distance functions lose their usefulness as the dimension of the space increases (Beyer et al. 1999). We therefore employ a common dimensionality reduction process known as Singular Value Decomposition (SVD), a generalization of Primary Component Analysis, as a means to reduce the dimensionality of our data. We note that the zone diversity metric introduced above can be computed either for the unreduced data or the reduced data—the definition holds for both perspectives.

A common technique in recommendation algorithms and dimensionality reduction, SVD allows the reduction in size of the data while still capturing valuable features (Leskovec et al. 2014). This dimensionality reduction is done by projecting the data with a set of orthogonal basis vectors representing the modes of variance in the system. These vectors are often referred to as the "concept space," as each vector represents some abstract concept which captures the variance. Truncating the lower variance orthogonal vectors before reconstruction yields a best approximation of the data in lower dimensional space, and we can select the number of dimensions depending on how much information we would like to retain during reconstruction. We apply SVD to a transposed version of our activity states: $\mathbf{M} = \mathbf{S}^T$ (where $\mathbf{M}$ has $D \times T$ time rows and I occupant columns) Here, the number of rows is expected to be much larger than the number of columns. The decomposition is as follows:

$$\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{5.2}$$

Here, $\mathbf{U}$ contains the eigenvectors of $\mathbf{M}\mathbf{M}^T$, $\mathbf{V}$ contains the eigenvectors of $\mathbf{M}^T\mathbf{M}$, and $\mathbf{\Sigma}$ is a diagonal matrix containing the singular values of $\mathbf{M}$ ($\mathbf{\Sigma^2}$ contains the eigenvalues of $\mathbf{M}^T\mathbf{M}$). The shapes of these matrices are determined by $r = \text{rank}(\mathbf{M})$, where in our case, $r = I$. Therefore, $\mathbf{U}$ has the shape $(D \cdot T) \times I$. These eigenvector matrices can be thought of as a rigid transformation in high dimensional space, which aligns the data according to the variance of the data. Thus the primary axis after the rotation will be aligned with the axis of highest variance, the second axis will be aligned with the second highest variance, etc. In practice, the matrices $\mathbf{U}$ and $\mathbf{V}$ map the data to a concept space. The concept space is defined by the shape of the primary eigenvectors in the system, which typically will provide some kind of intuition as to what is driving the variance. In occupant schedules, a concept space might identify the time at which a person usually arrives at the office to be a valuable indicator, or when they take their lunch break.

Our purpose for using SVD is to project very high dimensional occupant behavioral space to a much lower dimensional representation, permitting a richer clustering process. After the matrix is decomposed into $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{\Sigma}$, the original data matrix $\mathbf{M}$ can be projected into concept space via a rigid rotation from $\mathbf{U}$: $\mathbf{R} = \mathbf{U}^T\mathbf{M}$. The result of this projection is the condensing of the state data into an $I \times I$ matrix $\mathbf{R}$, where each column, previously the length of the full time series, is now

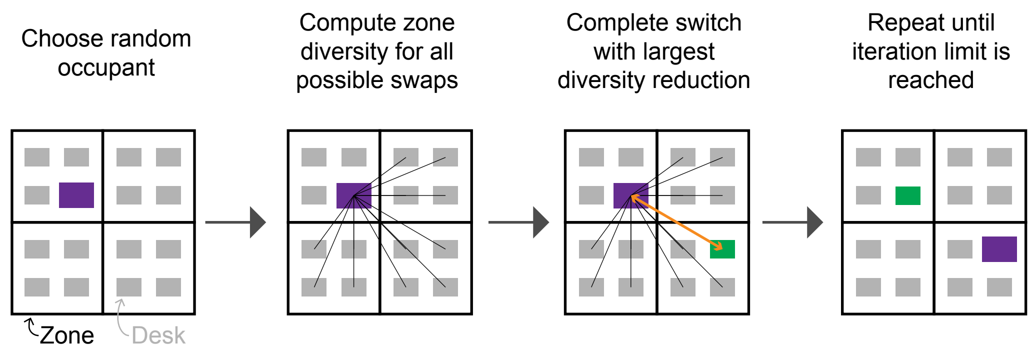

Figure 5.2: Occupant clustering algorithm.

more densely represented, and the different columns correspond to the different occupants. Because is orthogonal, the removal of the least powerful vectors prior to multiplication with $\mathbf{M}$ yields a projection of the data into the lower dimensional space with the least amount of lost value. For example, for a d-dimensional representation of the data, such that $d + d' = I$, the least significant d' eigenvectors/values in $\mathbf{U}$ are removed and the resulting representation of the full dataset $\mathbf{R}'$ will have the dimension $d \times I$. This approximates the replacement of the smallest eigenvalues with a 0 term, which becomes redundant in the reconstruction and can be truncated without losing value.

Stochastic constrained Expectation Maximization occupant clustering

With this relatively low-dimensional representation of our activity data $\mathbf{S}$, data-driven clustering of the occupants according to their activity states becomes more feasible. Here, we introduce a novel clustering algorithm based on the data representation $\mathbf{R}'$. The objective of the clustering algorithm is to minimize the zone diversity metric from section 5.2.1 for the occupants within a building zone. Our problem setting has the real-world constraint that each of the building zones retain their same size at the end of the clustering routine to preserve the same overall occupant spatial density, preventing the use of standard clustering algorithms such as k-means. The intuition behind our novel approach is to minimize the zone diversity metric by spatially swapping occupants with other occupants that reduce zone diversity. By doing so, we can expect to reduce lighting consumption according to the relationship we have established between these metrics.

The mechanics of the algorithm are depicted in Figure 5.2. First, we choose a random occupant/desk, which is associated with a building zone. We note that not all building zones need to be the same size as they are in the simple depiction of the algorithm in Figure 5.2. We then simulate a "swap" between this occupant and all occupants in the other zones of the building. The resulting swap will be the shift which had the greatest global drop in zone diversity, which includes the null action of the node swapping with itself. We repeat this process until an iteration limit is reached.

5.2.3 Optimizing layouts: Genetic algorithm

Our problem statement, to optimize building layout in order to reduce the energy consumption of building systems, has an extremely large possible solution space. If there are I occupants in a building and n possible zones to assign them to (of equal size $m = I/n$), then the number of possible assignments can be computed as follows:

$$\#\text{ of possible groupings} = \frac{I!}{(m!)^n \cdot n!} \tag{5.3}$$

For example, with 50 occupants and 5 groups, the number of possible assignments is on the order of 10^{29}. In addition to the large solution space associated with our problem, the effects of reassigning occupants are expected to be highly nonlinear. When optimizing in these circumstances, genetic algorithms have been shown to perform well (Nagy et al. 2017; Razavialavi & Abourizk 2017; Sonta & Jain 2020). We therefore implemented a custom genetic algorithm to assign occupants to desks and optimize building layouts, as described below.

Genetic algorithms belong to the class of evolutionary algorithms for optimization, originally inspired by the process of natural selection. The process begins by creating an initial set of design points—in our case, building layouts. Each building layout x in the initial population $\mathbf{P}$ is defined by the grouping of occupants to the zones of a building. A fitness function is used to evaluate the fitness of each design point $f(x)$. For this fitness function, we leverage a data-driven surrogate simulation engine that can be used to predict building energy consumption based on occupant schedules and other time series information, as discussed below in section 5.2.4. Once each design point is evaluated, a certain number of designs are selected to create a new generation of designs. In our case, we select the $\mathbf{B}$ best performing layouts, and, in order to maintain diversity in the population, we also select $\mathbf{R}$ random layouts. Among the best performers and randomly selected layouts, two designs are chosen at a time and recombined c times to form the designs in the next generation. The first step is crossover, whereby for each desk location in each zone, the occupant selected to occupy that desk is a random selection of the two occupants in the original two layouts. The next step is mutation, which occurs for each new individual with probability m. If mutation does occur, a random desk in each zone is swapped with a random desk from a random other zone. Crossover is meant to preserve the high-performing features that exist in the best-performing designs in the previous generation; mutation is meant to introduce randomness so that the algorithm does not get stuck in a local minimum. Once a completely new generation is created from the previous generation, through crossover and mutation, the process repeats for G generations. The parameters, therefore, that must be chosen to run the genetic algorithm are the fitness function f, the population size $|\mathbf{P}|$, the number of best performing layouts $|\mathbf{B}|$ and the number of randomly chosen layouts $|\mathbf{R}|$, the number of new layouts c created for each chosen pair, the mutation probability m, and the number of generations G. Figure 5.3 shows a visual representation of the algorithm.

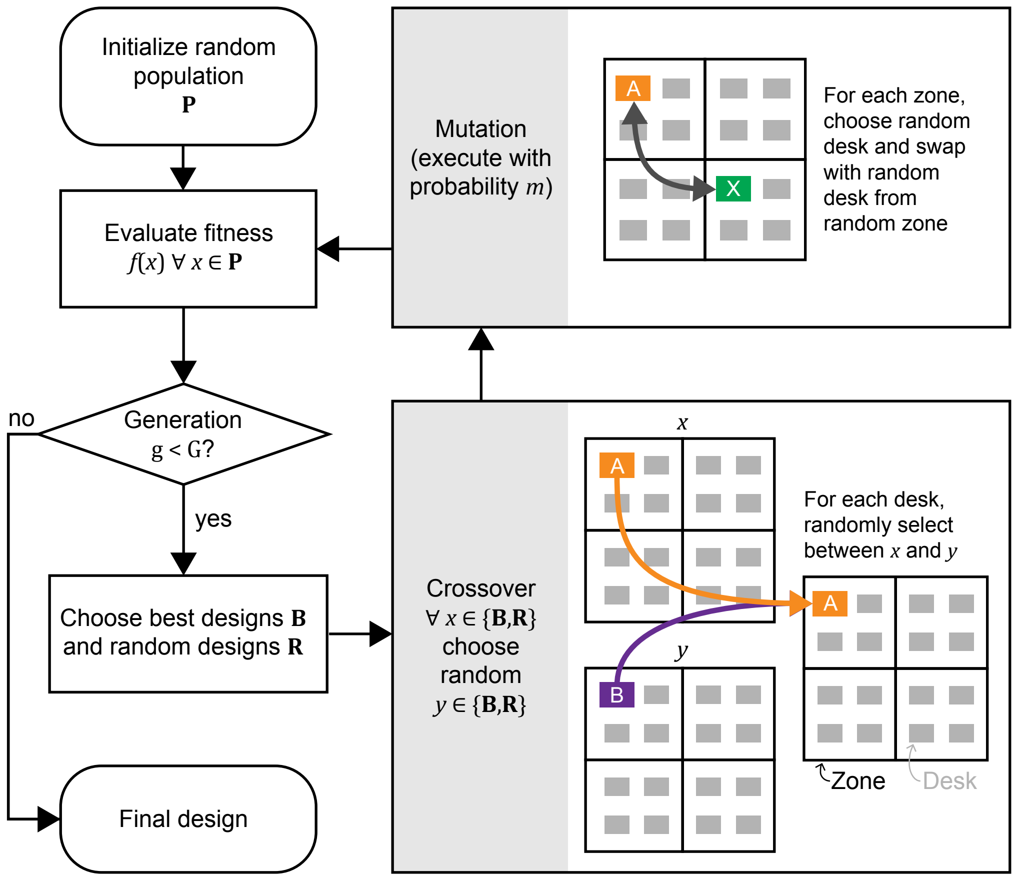

Figure 5.3: Genetic algorithm adapted for building layout optimization.

5.2.4 Simulating lighting energy consumption based on building layouts

There are several viable simulation models for evaluating the expected energy consumption of different building layouts. These models fall into two main categories: physics-based thermodynamic models (e.g., EnergyPlus (U.S. Department of Energy n.d.)) and data-driven "surrogate" models (Aijazi & Glicksman 2016). Thermodynamic models have been shown to be particularly helpful when evaluating energy consumption from heating and cooling systems, though recently mixed models and data-driven surrogate models have become more prevalent. However, because our analysis focuses on lighting systems with direct control through occupancy sensors (as explained below), the thermodynamic models are more complex than necessary for this task. Additionally, the time required for their analysis is prohibitive for running our genetic algorithm optimization. Therefore, we sought to develop a data-driven surrogate model that utilizes machine learning to predict lighting energy consumption based on occupant schedules as well as standard time series features (e.g., time of day,

day of week, etc.). We chose to test multiple linear regression (MLR), support vector regression (SVR), random forests (RF), and artificial neural networks (ANN) to determine the most robust surrogate model for our purpose.

The 7 specific features we use for prediction of energy consumption are as follows:

- s_1, s_2, s_3: number of occupants in each of the three energy states as defined above in section 2.3.
- Hour of day (0–23)
- Day of week (0–6)
- Weekday/weekend indicator (0 or 1)
- Zone number (0–number of zones)

The zone number is included to enable the model to adapt to zone-specific operational tendencies. For example, many lighting systems include daylight sensors, whereby the artificial lighting levels are lowered if enough daylight is present. This would be expected to vary throughout the day for each zone, depending on orientation and other factors.

We implemented each model, described briefly here, using the scikit-learn package in Python (Pedregosa et al. 2011).

- *Multiple linear regression.* The simplest model, MLR seeks to predict energy consumption ($\hat{\mathbf{y}}$) as a linear combination of the model features ($\mathbf{X}$): $\hat{\mathbf{y}} = \beta\mathbf{X} + \epsilon$, where β is a vector of parameters and ϵ is the error term. The fitting of the parameters involves minimization of the error term.

- *Support vector regression.* The model produced by SVR relies on a small subset of the training data known as support vectors. Errors within the bounds created by the support vectors (within margin ϵ) are ignored. Fitting an SVR model involves the following optimization:

$$\begin{aligned} &\min_{w,b,\zeta,\zeta^*} \frac{1}{2} w^T w + C \sum_{i=1}^{n} (\zeta_i + \zeta_i^*) \\ &\text{subject to } y_i - w^T \phi(x_i) - b \le \varepsilon + \zeta_i, \\ &\qquad w^T \phi(x_i) + b - y_i \le \varepsilon + \zeta_i^*, \\ &\qquad \zeta_i, \zeta_i^* \ge 0, i = 1, ..., n \end{aligned} \tag{5.4}$$

  where w is the set of feature weights, ζ_i and ζ_i^* are the residuals beyond ϵ, and ϕ is a kernel function that is often nonlinear such as the Gaussian radial basis function. The SVR model has been previously applied to building energy prediction tasks with success (Jain et al. 2014).

- *Random forests.* The RF regression model is an extension of the decision tree model in regression form, in which the overall model aggregates (generally through averaging) the result from many independently fit trees. Each tree constitutes a series of decisions on the features (e.g., time of day is less than or greater than 6 a.m.), and once the full series of decisions are made, a final value is chosen. RF models have successfully been applied to energy prediction in various settings (Ahmad et al. 2017; Wang et al. 2018).

- *Artificial neural network.* An ANN is an interconnected group of nodes, in which each node produces a signal according to the data it receives. ANN architecture generally involves an input layer (which receives the features), one or more hidden layers, and an output layer (which produces a prediction). These networks are fit using backpropagation. They have been widely used for energy prediction tasks (Ahmad et al. 2017; Ekici & Aksoy 2009).

5.2.5 Empirical data

We installed plug load energy sensors at each desk on a single floor or a large commercial office building in Redwood City, CA. The floor comprises 165 desks, of which 151 are occupied. The data collection period started August 1, 2019 and ended February 29, 2020. The sensors are Zooz SmartPlugs that communicate to a Samsung SmartThings hub through Z-Wave technology. The plug load sensors reported power consumption values any time the power consumption varied by more than 0.1 W. Consistent with previous work (Chapter 2), we aggregate the power consumption to 15-minute intervals, as modeling space use at this time frequency has been shown to limit noise while providing useful information in terms of building operation. As described above in section 2.3, we map the raw plug load data to energy states describing occupant schedules.

The office building is equipped with a lighting system that operates based on occupancy sensors, daylighting sensors, and schedules. The lighting zones are controlled by occupancy sensors across the building floor. If any lighting fixture within a zone senses motion over the past 20 minutes (10 minutes on weekends), all fixtures within that zone turn on. 11 of the lighting zones service all workspaces (the others service other small shared spaces such as meeting rooms, corridors, etc.) We restrict our analysis to the 11 zones that service workspaces, as we are interested in characterizing energy consumption in places where individual schedules can be modeled according to our data. The lighting energy data is available at 1-hour intervals for the full duration of the study.

Due to persistent sensor outages at the beginning of data collection, our analysis begins with data on October 1, 2019. In addition, due to a temporary shutdown of organizational activities over the New Year, we discarded data from December 16, 2019 to January 4, 2020. We therefore restrict our analysis to 132 full days of sensor and lighting energy data.

5.3 Results

5.3.1 Increased zone diversity correlates with increased energy consumption

For each day over the data collection period, and for each of the 11 lighting zones in the building, we compute the zone diversity, as discussed above in section 5.2.1. We also compute the average energy consumption across lighting fixtures within a zone. We then complete a regression analysis for the relationship between zone diversity and energy consumption, as shown in Table 5.1. We find that for each zone, there is a positive relationship between zone diversity and energy consumption. We compute the t-statistic for the regression coefficient, and we find that the p-value for the t-statistic is significant at the 0.001 level for all zones. While the positive relationship is clear, both the slope of the relationship and strength of that relationship in terms of the R^2 value vary across zones. In Figure 5.4, we show the data along with the regression lines and 95% confidence intervals for each zone. We note that this result serves as validation of this relationship between occupant schedules and energy consumption. In addition, it serves as motivation for optimizing building layouts in order to reduce this diversity and therefore save energy.

Table 5.1: Regression results.

Zone	Coefficient	(Standard Deviation)	t-statistic	p-value	R^2
0	422.5	(78.29)	5.396	< 0.001	0.165
1	492.4	(52.64)	9.354	< 0.001	0.354
2	828.2	(106.7)	7.762	< 0.001	0.288
3	532.4	(80.00)	6.655	< 0.001	0.288
4	527.9	(35.33)	14.944	< 0.001	0.627
5	749.6	(126.3)	5.934	< 0.001	0.253
6	514.3	(40.78)	12.611	< 0.001	0.288
7	989.2	(51.83)	19.087	< 0.001	0.666
8	177.6	(39.70)	4.472	< 0.001	0.117
9	486.9	(20.94)	23.256	< 0.001	0.757
10	97.1	(60.90)	4.878	< 0.001	0.157

5.3.2 Lighting energy simulation driven by occupant schedule data

We tested the four models described above in section 5.2.4 by splitting the data into a training set and a test set, and then performing 5-fold cross validation on the training set for model development. The initial split into training and test sets preserves the time series attribute of the data: the first 80% of the data, in terms of time, are used as the training set and the second 20% are used for testing. We preserved the time series attribute of the data in order to test the ability of the models to predict future events and also to enable time series visualization of the performance of each model.

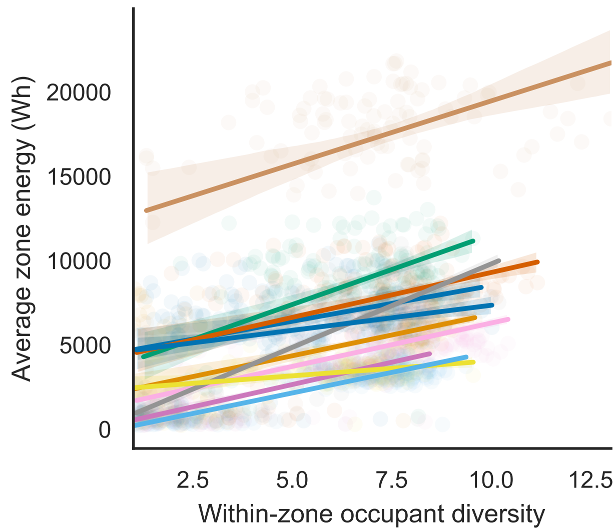

Figure 5.4: Relationship between zone diversity metric and energy consumption for each zone, along with regression fits and confidence intervals. Colors represent lighting zones.

For the MLR, ANN, and SVR (non-tree-based) models, one-hot encoding is used for the day of week and zone number features, bringing the total number of features up to 23 features. Additionally, for these models, the hour of day feature is decomposed using sine and cosine transformations, to preserve the cyclical nature of the hour features (i.e., hour 23 is close to 0). Furthermore, the "state count" features (number of occupants in each energy state) are transformed using a sigmoid function. The intuition behind this transformation is that there are diminishing effects of having more than one occupant present in the zone (as shown in Figure 5.5). This phenomenon occurs due to the fact that the lights only need to sense one occupant in order for all lights within a zone to turn on. Finally, the features are scaled to range between 0 and 1. Each of these steps were found to enhance the performance of the non-tree-based predictive models. These additional feature scaling steps are not used for the tree-based RF regressor, as the decisions are invariant to the scaling.

We find that, before tuning of hyperparameters, the ANN performs the best in terms of mean squared error (MSE) and explained variance (R^2), and the RF model performs the best in terms of mean absolute error (MAE) (Table 5.2). Because the squared error terms, MSE and R2, exaggerate the importance of larger errors, we can infer that the RF model occasionally produces larger errors, while the ANN model produces a higher baseline of error values. While the ANN model takes the longest time to train, it requires less computation time to produce a prediction than the RF model. Based on these results, we chose the ANN and RF models for additional hyperparameter tuning.

Our hyperparameter tuning for both models also involved 5-fold cross-validation on the training set. For the ANN model, we tuned the hidden layer sizes, activation function, solver, and learning

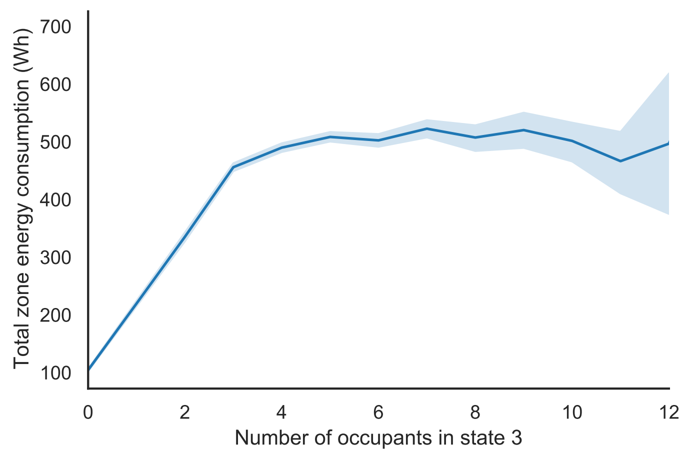

Figure 5.5: Total zone energy consumption vs. number of occupants in state 3.

Table 5.2: Prediction model results on 5-fold cross validation.

Model	Mean Absolute Error (MAE)	Mean Squared Error (MSE)	Explained Variance (R^2)	Time for Training (s)	Time for Prediction (s)
Multiple Linear Regression	9.55	141	0.534	0.0311	0.00198
Support Vector Regression	7.13	118	0.614	30.9	4.38
Random Forest Regression	6.11	98.2	0.678	2.82	0.0983
Artificial Neural Network	6.29	88.7	0.710	54.8	0.0105

rate. We performed a grid search on these hyperparameters and found the best parameters to be a single hidden layer of size 100, the tanh activation function, the Adam solver, and a learning rate of 0.01.

For the RF model, we tuned the number of trees, minimum number of samples to produce a split in the tree, the minimum number of samples per leaf, the maximum depth of the tree, and whether the bootstrap methodology was used in model training. We found the best parameters to be 200 trees, minimum split size of 50, minimum samples per leaf of 2, maximum depth of 300, and use of bootstrap in model training.

We found that our tuned RF model outperformed our tuned ANN, with an MAE of 6.27, MSE of 87.1, and R^2 of 0.715 as estimated through cross validation. We also note that while hyperparameter tuning did improve the performance of each model, in both cases the improvement was very small.

This high performance "out of the box" could enable wide adoption of similar surrogate models for the purposes of simulating energy consumption based on occupancy data.

Final results for the ANN and RF model, on both cross-validation and the test set, are shown in Table 5.3. Because the ultimate goal of our model is to be able to predict total energy consumption, as a result of features on the scale of 15 minutes, we also tested the performance of the ANN and RF models on a more aggregate level. We summed predicted and actual energy consumption to daily values and computed the R^2 metric on this aggregate level. As expected, both models benefit, in terms of their error rates, from this aggregation. However, the RF model benefits even further than the ANN. In Figure 5.6, we show the predicted energy consumption (using the tuned RF model) vs. actual energy consumption data for the first seven days of data in the test set for zone 1. As this time-series plot shows, the large "jumps" in energy consumption (i.e., from low values to high values and vice versa) generally match between the predicted and actual data, with these jumps likely driven by changes in occupancy. Much of the error seems to result from small differences rather than large mischaracterizations by the model.

Table 5.3: Prediction model results after hyperparameter tuning on both 5-fold cross-validation and final test set.

Model	Errors on CV			Errors on Test Set	
	Mean Absolute Error (MAE)	Mean Squared Error (MSE)	Explained Variance (R^2)	Explained Variance (R^2) Hourly	Explained Variance (R^2) Daily
Tuned Random Forest Regression	6.27	87.1	0.715	0.740	0.834
Tuned Artificial Neural Network	6.28	88.5	0.710	0.734	0.817

An additional strength of the RF model is its interpretability. Each split in a tree involves a decision on one of the features, which do not need to be scaled or one-hot encoded. As a result, the feature importances can be easily quantified and visualized. For the final RF model, we find the most important features to be time of day, and number of occupants in state 3 (Figure 5.7). This importance of the occupant feature underscores the notion that the lighting system operation is driven by occupant behavior, and therefore there exists opportunity to save energy by adapting the layout of the building to the behavior of occupants.

We note that the time for prediction using the RF model is roughly one order of magnitude above the time for prediction using the ANN model. In our case, these times were sufficiently small such that the difference between using them was inconsequential to our analysis. However, in situations with particularly high computing costs, the ANN model, while slightly less accurate, could be used to improve computational costs. That said, the RF model will always maintain a much higher level of interpretability.

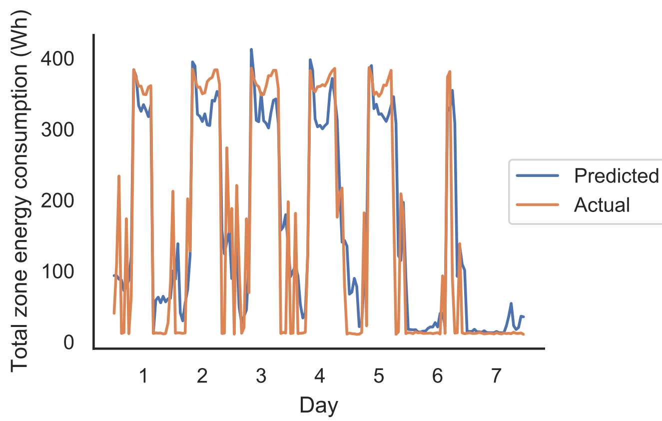

Figure 5.6: Example predicted (using tuned RF model) vs. actual energy consumption data for the first seven days of data in the test set for zone 1.

5.3.3 Optimizing layouts reduces energy consumption according to simulations

In this section, we show the results of our clustering-based and genetic algorithm optimization methods on two examples: a synthetic example that demonstrates the underlying mechanics of the optimization routines and the surrogate simulation model, as well as the empirical data collected from the office building in Redwood City.

Synthetic example

Here, we introduce a simple example of an office building based on the hypothetical example from the introduction, though updated to be more representative of common office behavior. Our purpose in showing this synthetic example is to demonstrate the mechanics of the optimization methods in creating new building layouts based on occupant activities. In this synthetic example, there are four different "archetypes" of occupant schedules. Each archetype defines the following behaviors:

- **Arrival.** Time when occupant arrives at the building, transitioning from low energy state to a medium or high energy state.
- **Lunch.** Time and length of occupant's lunch, in which the occupant transitions to a low energy state for the duration of lunch.
- **Meeting(s).** Time(s) and length(s) of the occupant's meeting(s), in which the occupant transitions to a low energy state for the duration of the meeting(s).
- **Departure.** Time when occupant leaves the building, transitioning to a low energy state.

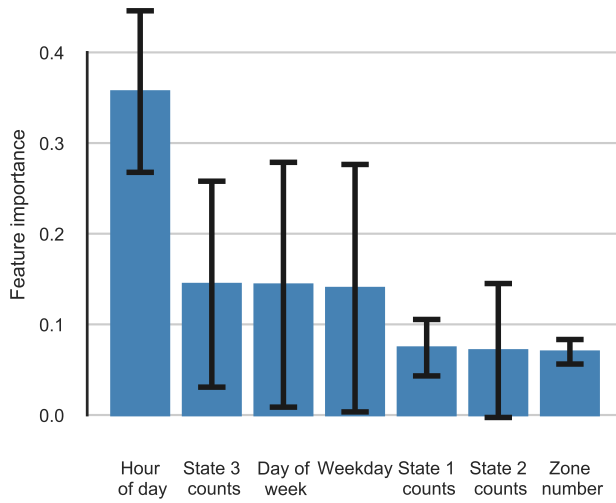

Figure 5.7: Feature importance for final random forest regression model.

When the occupant is in a normal working state, between arrival and departure but not during lunch or a meeting, the occupant randomly transitions between high and medium energy states. This behavior models normal occupant behavior of taking short breaks throughout the workday. The four specific archetypes used in this synthetic example are shown in Table 5.4.

Table 5.4: Archetypes used in the synthetic example.

Archetype	Arrival	Lunch	Meetings	Departure
1	9am	12pm (1 hour)	3pm (1 hour)	5pm
2	9am	None	None	4 pm
3	11am	3pm (1 hour)	3pm (1 hour)	7 pm
4	7am	11pm (1 hour)	1 pm (2 hours)	5 pm

The hypothetical building includes four rooms, with nine desks per room. Each occupant follows one of the four archetypes and there are exactly 9 of each archetype. We simulate energy consumption for one day, leveraging the random forest model discussed above in section 5.2.4. For simplicity, we assign each room to be the same zone number (0), which in the empirical example, has roughly the same size of 9 occupants. We also assume the day is a Monday. We test two different layouts to show the effects of layout on energy consumption: a random layout, and a layout in which all 9 occupants of each archetype share a room. We find, leveraging the lighting energy surrogate simulation model trained on real-world data, that the archetype-based layout creates a situation in which the lighting

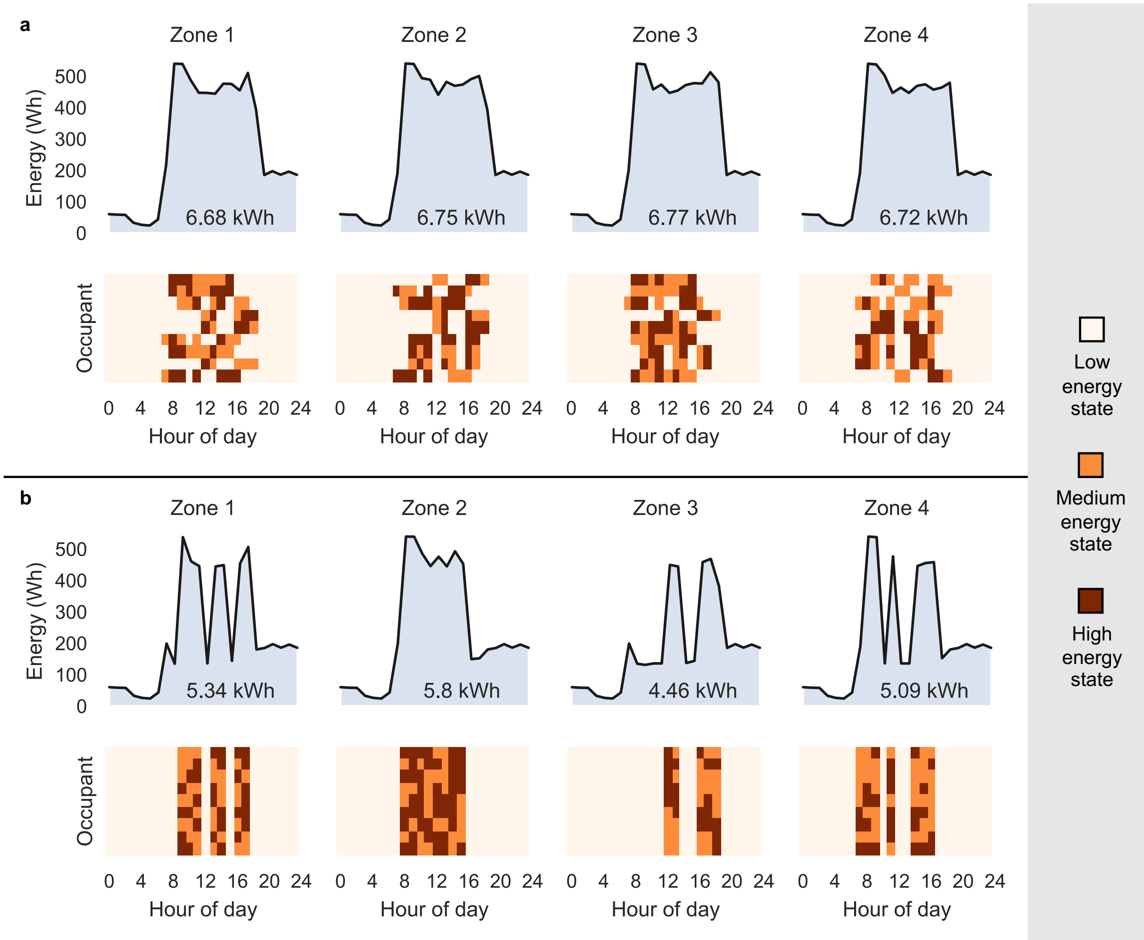

Figure 5.8: Synthetic example occupant schedules and simulated energy consumption for (a) random layout and (b) known optimal layout.

system is able to respond to occupants' use of the space (Figure 5.8). The result is significant energy savings of 22%.

We know the layout in which the building is arranged by archetype is optimal by inspection. This *a priori* knowledge of the system enables evaluation of the optimization routines introduced in sections 5.2.2 and 5.2.3. When we apply the clustering-based optimizer to this example, we find that the optimizer is able to arrive at the known optimal layout very quickly (Figure 5.9). The genetic algorithm, on the other hand, quickly finds a near-optimal solution, but is unable to reach the fully optimal layout. This behavior of genetic algorithms—in which they come close to optimality but have final convergence issues—is well documented. However, this synthetic example is a relatively simple example. In a real-world office building, with much more variation in occupant activities

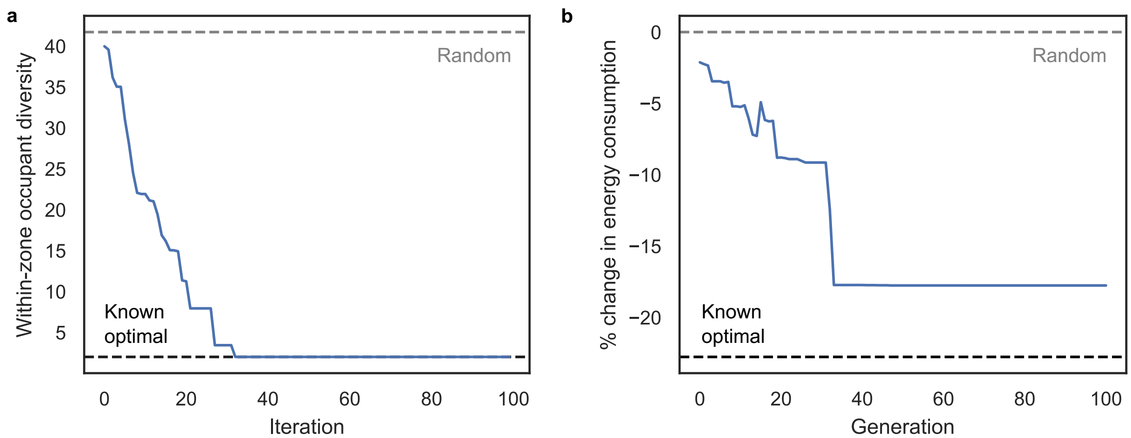

Figure 5.9: Optimization convergence on synthetic example, in relation to a random layout and the known optimal layout for (a) clustering-based optimization and (b) genetic algorithm. The same optimal layout produces the zone diversity value and energy values shown in the figure.

both in time and space, we can expect the genetic algorithm to do well in avoiding the many local minima that may exist.

Empirical case study

The synthetic example offers insight into the mechanics of the optimizers developed in this work, but the overly simplistic nature of that example limits claims that may be made about the extensibility of our framework. We therefore leverage the optimization algorithms and surrogate simulation model to optimize the layout of the Redwood City office building introduced in section 5.2.5. We simulate energy consumption using the full dataset of 132 days of occupant behavior. We start by estimating the energy consumption of the lighting system from 100 random building layouts as well as the existing building layout. We then apply the clustering-based optimization routine to the data with varying dimensionality (3, 5, 10, 30, 100, 151, and full dimensionality without reduction), as well as the genetic algorithm. We run each optimizer 100 times to create 100 layouts and ultimately a distribution of estimated energy consumption (Figure 5.10). Overall, we find that the existing layout performs slightly better than the random layouts, but that further improvements on energy consumption can be realized by optimizing the layouts for energy efficiency. Increasing the number of dimensions used in the clustering optimizer improves performance.

We also computed the zone diversity in occupant schedules as introduced in section 5.2.1. Here, the zone diversity metric is computed for the entire 132 days, as opposed to daily in section 5.2.1. In general, we find that the zone diversity metric follows the predicted energy consumption (Figure 5.11). However, we note that while the layouts produced through the genetic algorithm are among

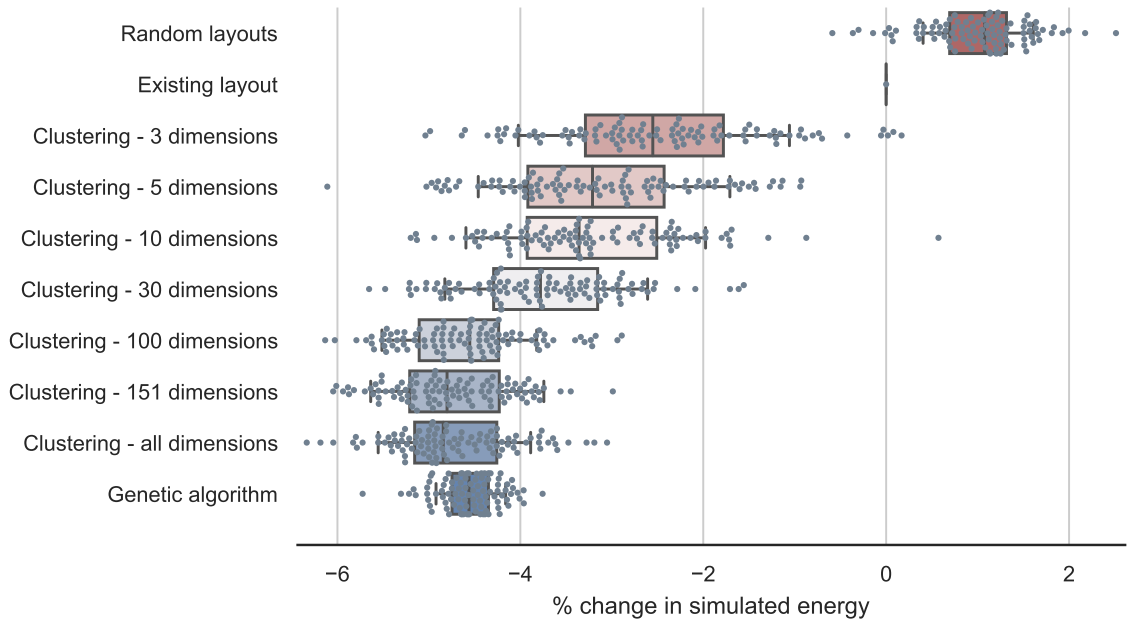

Figure 5.10: Simulated energy consumption (expressed as % change from the existing layout) for random and optimized building layouts.

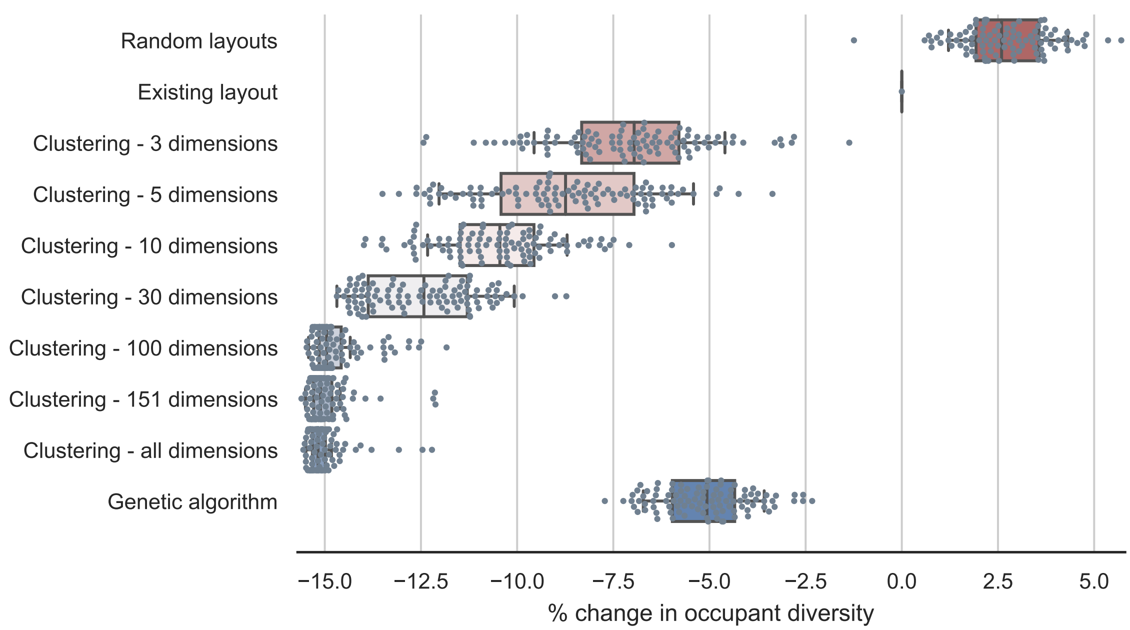

Figure 5.11: Change in zone diversity metric (expressed as % change from existing layout) for random and optimized building layouts.

the best, the zone diversity metrics for the genetic algorithm layouts are substantially higher than for the clustering-based layouts.

5.4 Discussion

Our results indicate the possibility for significant energy savings of the lighting system through redistribution of occupants in space. This energy savings is possible by leveraging the occupancy-sensing feature of the lighting system. We note that while the heating, cooling, and ventilation system was not modeled in our case study due to data limitations, we would expect to find similar results as long as the thermal system is able to reduce service during periods without occupancy. The analysis of expected effects on these additional building systems is a rich area for further work.

While optimization creates reductions in energy consumption compared to the existing layout, it is interesting to note that the existing layout has a smaller simulated energy consumption than 90% of the random layouts. This is not altogether surprising, as it indicates that behavior in the existing building is more aligned than if each occupant behaved independently. This result may suggest that people tend to adapt to the behavior of those around them, as has been documented in previous social science research (Chartrand & Bargh 1999). This question of how people adapt their behavior to their surroundings has potentially large implications for interpretations of our results. Underlying our analysis is the assumption that individual behavior will not change when occupant layouts are changed. We note that this assumption is a limitation of our model, and that behavior can be expected to change in some way in response to new desk assignment changes. The direction of the impact of these behavioral changes on energy consumption is unclear. However, our finding that the existing layout is more efficient than a random layout suggests that people adapt their behavior to match those around them. It is therefore possible that when we create layouts in order to reduce within-zone diversity in behavior, occupants may naturally choose to adapt their behavior and therefore further reduce within-zone diversity. On the other hand, it may be possible that occupants may choose to alter their behavior in other ways, either by attempting to maintain social behaviors that precipitated out of their previous setting, or perhaps by choosing to separate themselves from the others in their new surroundings. We note that the evidence from this paper seems to suggest that individuals, at least to some extent, tend to assimilate their behavior to those around them.

A key finding from this work is that the reduction in energy consumption from both the clustering-based optimization (when enough dimensions are used) and the genetic algorithm is roughly the same (Figure 5.10). This is particularly interesting because the clustering-based optimization procedure does not include any information about the surrogate model used for predicting energy consumption, whereas the genetic algorithm uses that model as feedback during its execution. We argue that one can reasonably view the genetic algorithm as closer to a “best-case scenario” for predicted energy consumption, given its explicit use of simulation as a feedback mechanism during optimization. Our finding that clustering, once enough dimensions are used to represent the data, performs just as well demonstrates the strength of the nave clustering-based approach. When only a small number of dimensions are used for clustering, the optimization does result in energy reduction, though the effect is smaller. It is therefore important to ensure that enough dimensions are used for representation of the occupancy data.

The effect of clustering on the within-zone occupant diversity metric is clear: substantial diversity can be reduced by clustering occupants according to their schedules. The genetic algorithm also reduces this zone diversity, but to a significantly smaller degree (Figure 5.11). This finding suggests that there are other factors beyond zone diversity that are addressed during execution of the genetic algorithm. We unfortunately cannot interpret what these factors are, but they could involve unique aspects of the lighting system's operation. It is also possible that the genetic algorithm is optimizing for uncertainty in the random forest surrogate simulation model, which could be a limitation of that approach.

We found that the clustering-based optimization works about as well as the genetic algorithm. The genetic algorithm is only executable when a simulation engine is available, which makes the clustering-based approach more practical in situations when only occupancy data are available. We note, however, that if a simulation model is available, the genetic algorithm can be seeded with the layouts obtained through clustering. In our case study, the clustering-based layouts form a distribution with regard to simulated energy consumption (Figure 5.10). In a preliminary analysis, we found that seeding the genetic algorithm with 50 clustering-based layouts and 50 random layouts created new layouts that performed as well as the best layouts from the clustering (the layouts furthest to the left on the distribution in Figure 5.10). Therefore, this ensemble approach may be useful in reducing uncertainty around the expected outcomes from either approach.

As discussed in section 1.3, optimization of building design and system control can create significant energy saving opportunities in buildings. For example, Krioukov et al. (2011) found that occupant-driven control of lighting systems can lead to 50–70% energy savings. These strategies are critical for making use of controllable building systems and achieving energy efficiency. In some ways, these strategies can be seen as the "low hanging fruit" of optimal building performance (e.g., turning the lights off when no occupants are present). The research we present is this paper takes these strategies one step further: our methodology is not just about optimizing building system operation using information about occupant dynamics, but also optimizing characteristics of those occupants—the layout of workstations—to take full advantage of building systems. The approach we introduce can be considered in tandem with the more traditional approaches of controlling energy-intensive building systems, and it has the added benefit of not requiring expensive upgrades to building automation systems in order to achieve additional energy savings. Furthermore, the approach we introduce can also be integrated with other characteristics of building design, such as thermal comfort or organizational success, both of which can be tied to building layout (Nagarathinam et al. 2018; Peponis et al. 2007).

The ultimate goal of the design optimization in this paper is to reduce energy consumption, a goal that is important for reaching our sustainable energy goals and reducing costs for organizations. However, it is essential to note that there are other goals for the success of the building and organization that should be considered as well. Chief among these goals is productivity—perhaps the ultimate purpose of commercial office buildings—which is difficult to define and varies among different organization. Numerous factors have been shown to influence productivity, including occupant thermal comfort (Keeling et al. 2012), organizational cohesion, and others. The benefit of the occupant-driven optimization-based design approach we introduce here is that there is a natural extension to include other objectives (such as those discussed above) through a multi-objective optimization. We specifically note that to address organizational cohesion and collaboration, the approach of space syntax analysis (Bafna 2003; Peponis et al. 2007) may yield viable optimization objectives. Other graph theoretical approaches that are designed to leverage organizational needs (Lather et al. 2019) could also complement the approach we introduce here. The integration of occupant-centric design for energy with design for organizational outcomes will be a rich area of future research investigations.

5.5 Conclusion

In this paper, we explored the relationship between occupant behavioral dynamics and energy consumption from energy-intensive building systems. We introduced (1) a zone diversity metric adapted from the literature for comparison with empirical building energy data, (2) a clustering-based building layout optimization methodology made possible through dimensionality reduction, (3) a novel genetic algorithm for building layout optimization, and (4) a data-driven surrogate simulation engine for predicting lighting energy consumption from occupancy data. In a case study, we found a significant relationship between building occupant zone diversity and actual lighting energy consumption. We also found that our layout optimization methods can be expected to reduce lighting energy consumption by about 5% from the existing layout and 6% from a random layout. Overall, we show that reconsidering the design of layouts in existing buildings has significant potential for realizing energy savings. Additionally, the approach of changing layouts to achieve energy efficiency also enables the simultaneous consideration of other factors influence by building layout, including organizational performance and thermal comfort. These methods, when integrated with the many objectives that drive building management, will be critical to ensuring dynamic and lasting energy savings in buildings.

Chapter 6

Conclusion

The overarching purpose of the research presented in this dissertation was to develop novel methods and theoretical contributions that integrate social and environmental goals in the design and operation of commercial buildings. With a focus on understanding, modeling, and inferring the dynamics of building occupants using ambient sensor data, this work contributes to a paradigm of data-driven and occupant-centric building design and management.

6.1 Theoretical contributions

Chapter 2 describes a novel methodology for understanding building occupant activities at scale. We develop an unsupervised occupant sensing strategy that leverages plug load energy sensor data collected at the individual desk level. In a small validation study, we show that this strategy is at least as effective as other sensors specifically designed for occupant detection while minimizing privacy concerns. The added advantages are that the sensors are individualized, providing information with high granularity across time and space. Furthermore, these sensors provide additional information beyond simple presence and absence: using the plug load energy data as a proxy for activities, we categorize occupant behavior into *activity states*, which provide an abstracted but nuanced view of space use. These new ways of analyzing occupant dynamics are made possible through the novel integration of a data science perspective with explicit engineering knowledge on plug load energy systems. As a result of this approach, this work contributes a spatially and temporally granular occupant sensing approach with theoretical value for the computational civil engineering domain: demonstrating how useful information can be extracted from ambient sensing data. Furthermore, this work constitutes a contribution to the building management domain as a novel method for gaining additional insight into building operation. Finally, a key goal for this work was extensibility, which has been demonstrated by the fact that the method has been applied to multiple additional office settings (two of which are discussed in Chapters 3 and 5). Through the knowledge gained by its mechanics and extensibility, this method has enabled novel socio-organizational and energy-focused research as discussed throughout this dissertation.

Chapter 3 extends the methodological approach introduced in Chapter 2 to the realm of organizational behavior. We introduce a novel method, the Interaction Model, for inferring socio-organizational network structure based on ambient sensing data. Leveraging activity states, as inferred from plug load energy data collected at the desk level, we further analyze occupant space use through a social lens. Our key insight was defining *opportunities for social interaction*, which can be interpreted as times throughout the day when two occupants have a high possibility of interacting in the building. We show that a network inferred through this perspective correlates—to a statistically significant degree—with network data collected through surveys. Two additional key findings point to the power of this method. First, we find that only 10 weeks of data are required for network inference, which is useful amidst the reality of dynamically changing spaces, organizations, and networks. Second, we show through a case study that the socio-organizational network does not correlate with the spatial layout of the building, which points to the possibility for redesigning the space in a way that matches the organizational structure of the occupants, thereby creating more opportunities for collaboration. Overall, the Interaction Model is a contribution to the domain of organizational behavior as a novel approach for measuring the structure of socio-organizational relationships. This work also has direct implications for the field of architecture, as it provides new information, obtained through ambient sensors, about the users of created structures. Through the comparison and continual analysis of spatial networks (e.g., space syntax) in relation to the inferred socio-organizational network, designers may be able to make decisions that enable greater connection between the spatial design of commercial buildings and the operation of the organizations within them.

Chapters 2 and 3 introduce novel methods for inferring useful information about building occupants. In Chapter 4, we adapt a novel method from the signal processing literature for analyzing building occupant dynamics in the context of social, spatial, and temporal data. Our Occupant Energy Signal Processing on Graphs ($\mathrm{OESP_G}$) framework continuously analyzes time-series data on occupant behavior ascribed to a graph defining social and spatial relationships. We show that this contextualized analysis reveals instances in which occupant behavior is *out-of-sync* with the expectations embedded in the combined spatial and social graph. This *out-of-sync* analysis enables identification of possible realignment strategies that increase overall alignment with building operation. The work in this chapter contributes to the computational civil engineering literature a new strategy, inspired by the field of signal processing, for analyzing building operation and energy consumption using ambient plug load sensors. The finding that altering the layout of occupants can improve the alignment between behavior and building operation also creates actionable opportunities for the fields of architecture and building management.

Chapter 4 introduces a new framework for analyzing occupant behavior, which Chapter 5 builds upon by introducing a framework for the *optimization* of building layouts. In a case study, we analyze occupant activity patterns (using the methodology introduced in Chapter 2) alongside an energy-intensive building lighting system. We find that with higher *diversity* in building occupant schedules within individual zones, controllable lighting systems consume more energy. We additionally introduce a novel data-driven energy simulation method that predicts lighting energy consumption using only occupant activity states and standard time-series information. After clustering occupants according to their activity patterns, we find through this simulation that lighting energy consumption can be expected to decrease. This optimization of building layouts shows that occupant-centric building design has the potential to improve operation of building systems. The integrated approach taken in this chapter demonstrates the opportunities for blending computational approaches (i.e., data science, machine learning, and optimization) with the architecture and building engineering disciplines. In the research context presented in Chapter 5, this dual approach can create new opportunities for energy savings in buildings. When considered in the context of the other chapters in this dissertation—particularly Chapter 3—these findings also contribute toward an overarching research framework that can integrate our energy-focused goals with socio-organizational goals such as the success of organizations.

6.2 Practical contributions

The methodological contributions discussed above were designed with practical application potential in mind.

- The occupant activity state inference methodology from Chapter 2 has been applied to 5 separate settings, ranging in size from 7 to 165 occupants, without requiring any parameter tuning or adjusting of the underlying mechanisms. This extensibility demonstrates the ability for this method to be deployed in practice, which can enable real-time feedback of occupant space use for building managers. The underlying sensing strategy required for this method—plug load energy sensors—are becoming prevalent and are often required for new construction for monitoring and control of plug loads (California Energy Commission 2018). When they are not already installed, they are relatively inexpensive and gaining traction the market as evidenced by new companies in the area (e.g., Keewi). This near ubiquity can enable wide adoption of real-time occupant activity sensing.

- Inferring the socio-organizational network of building occupants (Chapter 3) can enable greater adoption of social network analysis in building management. The data-driven Interaction Model eliminates the need to run expensive surveys or observational studies, creating the possibility to embed social information in building and organization management decisions. This social network information is critical to managing companies focused on innovation and collaboration (Cross et al. 2002).

- The analysis and optimization frameworks introduced in Chapters 4 and 5 can be used in practice to evaluate current building layouts and optimize them through rearrangement. These frameworks complete the loop of a “building occupant audit” enabled through simple ambient sensors, in which building design can be reconsidered through the context of occupant behavioral dynamics. These new approaches are becoming increasingly important as the uptake of “smart building” technologies continues to increase, and as we continue retrofitting our commercial buildings (Grimard 2016).

6.3 Privacy and ethical considerations

Much of the analysis conducted in this dissertation is built on data collected from ambient plug load energy sensors. These sensors on their own are common—and growing in prevalence—as tools for gathering information about building operation. However, this research shows that additional information beyond physical attributes can be captured through data inference—including occupant use of building spaces and social and organizational ties. This mining of information about human systems introduces concerns about privacy that should be carefully considered during research phases as well as any possible deployments in practice. Below, I discuss through a data-privacy lens both the strengths of the proposed methods as well as important considerations that arise from the research.

The sensors used in this study have an inherent advantage over other kinds of sensors that might be used to study occupant behavioral dynamics in buildings. Sensors such as video cameras capture far more information that might be considered an invasion of privacy when compared to ambient sensors such as plug load energy sensors. If one is concerned with the question of when and where occupants are using building spaces (the research objective of Chapter 2), the amount of information in a video or audio stream is much higher than this abstract notion of space use. Distilling the relevant space use information from the richer streams of video or audio data might be considered a form of "down-sampling" the data. On the other hand, inferring this information from plug load energy data required a novel transformation of the data using the techniques described in Chapter 2. One might consider this transformation as a form of "up-sampling" the data, since plug load data do not inherently provide information about occupant behavior. The abstract nature of the plug load energy data reduces the possibility of using the sensor data streams for any nefarious activities beyond the original intent. This ambient aspect of the research approach in this work helps to reduce such privacy concerns.

At the same time, it is important to be clear about what the methods in this dissertation enable: when analyzed properly, the sensing data can reveal information about human behavior and the structure of social systems. Therefore, any implementation of such systems should only occur when all parties are aware of the goals of the analysis (i.e., what the data are being used for). Furthermore, it is important for anyone deploying methods that capture personal information to follow emerging data privacy laws (e.g., the General Data Protection Regulation of the European Union, the California Consumer Privacy Act), which include the principles of transparency and anonymization, among others.

6.4 Suggested directions for future research

The research in this dissertation is in many ways a first step in creating an overall human-centric design paradigm for the built environment. The research methodologies introduced pave the way for new studies and the results introduce new questions that build upon my findings. I describe some avenues for further research below.

6.4.1 Occupant-centric building design and management

As we gain a deeper understanding of occupant behavioral dynamics in buildings, many questions related to optimal design and management of buildings become possible to investigate. Chapter 5 discusses an optimization framework for reducing building energy consumption by rearranging occupants in space. This redesign, while clearly helpful from an energy perspective, may have unintended consequences on the organizational dynamics within the buildings. Fortunately, the methods in Chapter 3 enable inference of the true organizational structure. Future design optimization methodologies can explicitly incorporate the occupant network structure as an optimization objective: What building layouts improve energy efficiency *and* promote better collaboration in the building?

One important method that can be expanded upon, especially with regard to organizational dynamics, is the space syntax methodology discussed briefly in section 3.2.4. This way of representing building space can be investigated in more detail, particularly as it relates to the organizational network. Attempting to maximize the correlation between this spatial network and the socio-organizational network is a possible means to increase alignment between our buildings and the organizations they house.

An area of future work that will be important for extending this dissertation's contributions is studying the empirical effects of making perturbations to the building. For example: how do organizational dynamics actually change when we attempt to increase alignment between the spatial and social networks? Does collaboration increase as expected, or are there other factors that influence effective collaboration? Similar questions can be asked with regard to energy. How do true energy savings compare to expected energy savings as building layouts are changed? These field studies will be important for validating the theoretical contributions from Chapter 5.

Finally, there are other components of our human experience of buildings that can and should be incorporated into an integrated socio-environmental design framework. Occupant thermal and visual comfort is a complex objective that varies by outdoor environment and personal preferences. The rearrangement of building layouts offers an opportunity to explicitly address such comfort levels. For example, individual preferences can be incorporated into occupant groupings, whereby occupants with similar preferences share the same spaces.

6.4.2 Extensions to the community and urban scale

To realize an overall integrated human-centric and environmentally-focused framework for design of the built environment, future work may also consider extensions to larger spatial scales. Sensor data at the urban scale is revealing more information about the use of urban and public spaces, which will enable more detailed analysis of our human and social behavior contextualized to the built environment. Some general research questions that may address these opportunities are:

- What urban planning and urban design decisions create more walkable urban environments? This question addresses both environmental goals (reduced dependence on energy-intensive mobility) and social goals (increased health outcomes, possible increased social resiliency).

- How can communities be planned to reduce energy consumption of distributed energy resources while promoting social objectives? Especially as large organizations begin development of community-scale developments (e.g., Google), this question will be important for ensuring both energy efficiency and operational efficiency of the large organizations that will use these community spaces.

6.4.3 Concluding remarks

Our processes for designing and operating our built environment are incredibly complex, involving many decisions and influencing many systems. One of the goals of this dissertation was to demonstrate the need for a paradigm that integrates our social goals with our environmental goals in this design and operation process. With the global climate crisis continuing to grow in importance, we are in need of new ways to rethink the environmental impact of our built environment. I believe that one of the ways to promote the adoption of new environmentally sensitive technologies and strategies is to ensure that there are social benefits as well. This dissertation builds toward a framework for accomplishing this integration in commercial office buildings. However, there is much work to be done to promote both environmental and social sustainability across all scales of our built world.

Appendix A

Empirical plug load data and occupant behavior

This Appendix validates the notion that variations in typical office behaviors can have a substantial impact on metered plug load energy consumption. We utilized a HOBO Onset plug load logger to capture power draw and energy consumption at 20-minute intervals (the same interval as the plug load monitor used in Chapter 4) for a typical office set up, including a laptop charger, monitor, and coffee maker. Notes were kept during the 24-hour data collection period to understand how recorded behavior correlated with energy variations. Figure A.1 summarizes the findings. It clearly indicates how activities such as leaving the desk for a meeting or class can lead to highly noticeable changes in energy consumption at the workstation. The $\mathrm{OESP_G}$ framework introduced in Chapter 4 leverages these changes in its analysis.

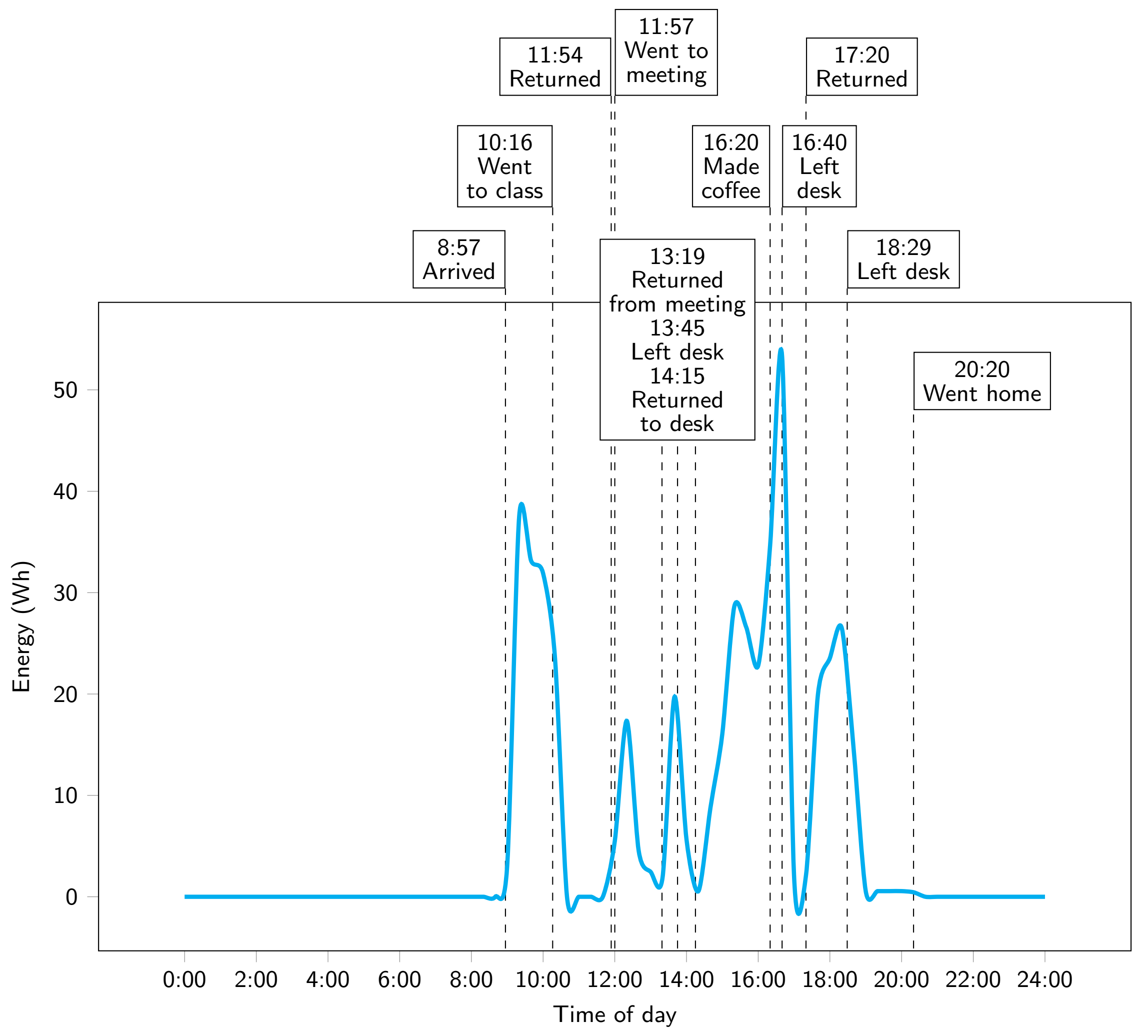

Figure A.1: Empirical plug load power data.

Appendix B

Survey for collecting social and organizational data

To benchmark all three inferred networks, we conducted a survey using questions from the social science literature that have been shown to measure social and organizational ties well. While no survey question is perfect, the questions used here were chosen due to their success demonstrated in other studies. At the beginning of the survey, respondents were asked to note with whom they had a personal relationship, a working relationship, or both from a list of all other occupants in the building. For each person identified, the respondents then were asked about both these personal (social) and working (organizational) relationships, as discussed below. The overall survey network is computed as an equally-weighted sum of the normalized social and organizational networks:

$$\mathbf{A}^* = \mathbf{A}^{*social_norm} + \mathbf{A}^{*org} \tag{B.1}$$

We measured social ties using the "inclusion of the other in the self" scale as discussed in (Gächter et al. 2015; Misra et al. 2016). This survey question makes use of the image shown in Figure B.1, which creates a Likert scale between 1 and 7, where larger values indicate closer personal relationships. A key benefit of this particular survey question is that it was designed to be easily understood very quickly and specifically to measure personal (social) relationships. Occupant i's response about occupant j becomes the entry $\mathbf{A}^{*social}_{i,j}$ in the social survey network. Because each occupant might interpret each value in the scale differently, we normalize the social survey network by row, where the normalized entry is calculated as follows:

$$\mathbf{A}^{*social_norm}_{i,j} = \frac{\mathbf{A}^{*social}_{i,j}}{max(\mathbf{A}_i)} \tag{B.2}$$

We measured organizational ties as first suggested in Krackhardt and Hanson's seminal paper about the complexity of organizations (Krackhardt & Hanson 1993). Specifically, for each occupant

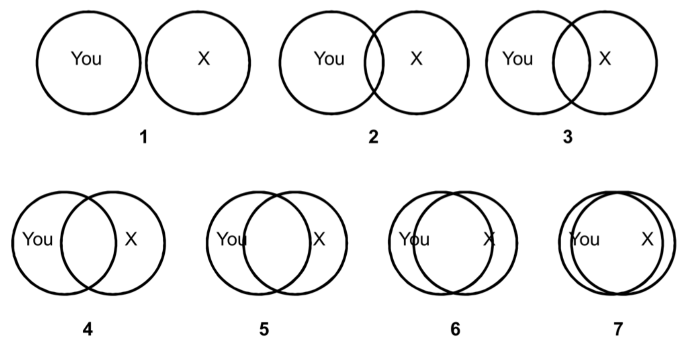

Figure B.1: Inclusion of the other in the self scale from (Gächter et al. 2015).

the respondent identified at the beginning of the survey, we asked the respondent the following three questions:

1. Communication: Have you received information at least twice in the last month from this person?

2. Advice: Have you received technical advice at least twice in the last month from this person?

3. Trust: Have you received personal work-related advice at least twice last year from this person?

To construct each component of the organizational network, occupant i's response about occupant j becomes the following entries: $\mathbf{A}^{*communication}_{i,j}$, $\mathbf{A}^{*advice}_{i,j}$, and $\mathbf{A}^{*trust}_{i,j}$. The overall survey network is then created as an equally weighted average of these three components:

$$\mathbf{A}^{*org} = \frac{\mathbf{A}^{*communication} + \mathbf{A}^{*advice} + \mathbf{A}^{*trust}}{3} \tag{B.3}$$

References

Agarwal, Y., Balaji, B., Gupta, R., Lyles, J., Wei, M., & Weng, T. (2010). Occupancy-driven energy management for smart building automation. In *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building* (pp. 1–6). Association for Computing Machinery.

Ahmad, M. W., Mourshed, M., & Rezgui, Y. (2017). Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and Buildings*, *147*, 77–89.

Ahmadi-Karvigh, S., Ghahramani, A., Becerik-Gerber, B., & Soibelman, L. (2018). Real-time activity recognition for energy efficiency in buildings. *Applied Energy*, *211*, 146–160.

Aijazi, A. N., & Glicksman, L. R. (2016). Comparison of Regression Techniques for Surrogate Models of Building Energy Performance. In *Proceedings of SimBuild 2016* (Vol. 6). Retrieved from https://ibpsa-usa.org/index.php/ibpusa/article/view/375

Akbas, R., Clevenger, C., & Haymaker, J. (2007). Temporal Visualization of Building Occupancy Phase. In *Proceedings of the International Workshop on Computing in Civil Engineering* (pp. 208–215). Reston, VA: American Society of Civil Engineers.

Akhavian, R., & Behzadan, A. H. (2015). Construction equipment activity recognition for simulation input modeling using mobile sensors and machine learning classifiers. *Advanced Engineering Informatics*, *29*(4), 867–877.

Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). A survey on sensor networks. *IEEE Communications Magazine*, *40*(8), 102–105.

Amayri, M., Ngo, Q.-D., & Ploix, S. (2016). Estimating occupancy from measurements and knowledge with Bayesian Networks. In *2016 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 508–513). IEEE.

Andersen, R. V., Toftum, J., Andersen, K. K., & Olesen, B. W. (2009). Survey of occupant behaviour and control of indoor environment in Danish dwellings. *Energy and Buildings*, *41*(1), 11–16.

Anderson, K., & Lee, S. (2016). An empirically grounded model for simulating normative energy use feedback interventions. *Applied Energy*, *173*, 272–282.

Anderson, K., Lee, S., & Menassa, C. C. (2014). Impact of Social Network Type and Structure on Modeling Normative Energy Use Behavior Interventions. *Journal of Computing in Civil Engineering*, *28*(1), 30–39.

Arora, A., Amayri, M., Badarla, V., Ploix, S., & Bandyopadhyay, S. (2015). Occupancy Esitmation Using Non Intrusive Sensors in Energy Efficient Buildings. In *Proceedings of the 14th Conference of International Building Performance Simulation Association* (pp. 1441–1448). Hyderabad, India.

Asavathiratham, C. (2001). *The influence model: a tractable representation for the dynamics of networked Markov chains* (Doctoral dissertation, Massachusetts Institute of Technology). Retrieved from https://dspace.mit.edu/handle/1721.1/33546

Azar, E., & Menassa, C. C. (2012). Agent-Based Modeling of Occupants and Their Impact on Energy Use in Commercial Buildings. *Journal of Computing in Civil Engineering*, *26*(4), 506–518.

Azar, E., & Menassa, C. C. (2014). Framework to Evaluate Energy-Saving Potential from Occupancy Interventions in Typical Commercial Buildings in the United States. *Journal of Computing in Civil Engineering*, *28*(1), 63–78.

Azar, E., & Menassa, C. C. (2016). Optimizing the performance of energy-intensive commercial buildings: Occupancy-focused data collection and analysis approach. *Journal of Computing in Civil Engineering*, *30*(5), C4015002.

Bafna, S. (2003). Space syntax: A brief introduction to its logic and analytical techniques. *Environment and Behavior*, *35*(1), 17–29.

Balaji, B., Xu, J., Nwokafor, A., Gupta, R., & Agarwal, Y. (2013). Sentinel: Occupancy Based HVAC Actuation Using Existing WiFi Infrastructure within Commercial Buildings. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems.* New York, NY, USA: Association for Computing Machinery.

Basbagill, J. P., Flager, F. L., & Lepech, M. (2014). A multi-objective feedback approach for evaluating sequential conceptual building design decisions. *Automation in Construction*, *45*, 136–150.

Beyer, K., Goldstein, J., Ramakrishnan, R., & Shaft, U. (1999). When is “nearest neighbor” meaningful? In *International conference on database theory* (pp. 217–235). Springer.

Bhatt, M., Hois, J., Kutz, O., & Dylla, F. (2010). Modelling Functional Requirements in Spatial Design. In *Proceedings of the 29th International Conference on Conceptual Modeling* (pp. 464–470). Berlin, Heidelberg: Springer-Verlag.

Bhatt, M., & Schultz, C. (2014). People-centered spatial design: On visuospatial cognition in the built environment. In *International Conference on Human Behavior in Design.* Ascona, Switzerland.

Bhatt, M., Schultz, C., & Minqian Huang. (2012). The shape of empty space: Human-centred cognitive foundations in computing for spatial design. In *2012 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)* (pp. 33–40). IEEE.

Bishop, C. M. (2006). Probability Distributions. In *Pattern Recognition and Machine Learning* (Vol. 4, pp. 67–127). Springer-Verlag New York.

Bonte, M., Thellier, F., & Lartigue, B. (2014). Impact of occupant's actions on energy building performance and thermal sensation. *Energy and Buildings*, *76*, 219–227.

California Energy Commission. (2018). *2019 Building Energy Efficiency Standards for Residential and Nonresidential Buildings.* California Energy Commission. Retrieved from https://ww2.energy.ca.gov/2018publications/CEC-400-2018-020/CEC-400-2018-020-CMF.pdf

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893–910.

Chen, J., Taylor, J. E., & Wei, H.-H. H. (2012). Modeling building occupant network energy consumption decision-making: The interplay between network structure and conservation. *Energy and Buildings*, *47*, 515–524.

Claudel, M., Massaro, E., Santi, P., Murray, F., & Ratti, C. (2017). An exploration of collaborative scientific production at MIT through spatial organization and institutional affiliation. *PLoS ONE*, *12*(6).

Clevenger, C. M., & Haymaker, J. (2006). The Impact Of The Building Occupant On Energy Modeling Simulations. In *Joint International Conference on Computing and Decision Making in Civil and Building Engineering* (pp. 1–10).

Costa, T., Boccignone, G., & Ferraro, M. (2012). Gaussian mixture model of heart rate variability. *PLoS ONE*, *7*(5), e37731.

Cross, R., Borgatti, S. P., & Parker, A. (2002). Making invisible work visible: Using social network analysis to support strategic collaboration. *California Management Review*, *44*(2), 25–46.

Deri, J. A., & Moura, J. M. F. (2015). Taxi data in New York city: A network perspective. In *2015 49th Asilomar Conference on Signals, Systems and Computers* (pp. 1829–1833).

Dijkstra, J., & Timmermans, H. (2002). Towards a multi-agent model for visualizing simulated user behavior to support the assessment of design performance. *Automation in Construction*, *11*(2), 135–145.

D'Oca, S., Chen, C. F., Hong, T., & Belafi, Z. (2017). Synthesizing building physics with social psychology: An interdisciplinary framework for context and occupant behavior in office buildings. *Energy Research and Social Science*, *34*, 240–251.

D'Oca, S., & Hong, T. (2015). Occupancy schedules learning process through a data mining framework. *Energy and Buildings*, *88*, 395–408.

D'Oca, S., Hong, T., & Langevin, J. (2018). The human dimensions of energy use in buildings: A review. *Renewable and Sustainable Energy Reviews*, *81*, 731–742.

Dong, W., Lepri, B., & Pentland, A. S. (2011). Modeling the co-evolution of behaviors and social relationships using mobile phone data. In *Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia* (pp. 134–143). New York, NY, USA: Association for Computing Machinery.

Doukas, H., Patlitzianas, K. D., Iatropoulos, K., & Psarras, J. (2007). Intelligent building energy management system using rule sets. *Building and Environment*, *42*(10), 3562–3569.

Dzeng, R.-J., Lin, C.-W., & Hsiao, F.-Y. (2014). Application of RFID tracking to the optimization of function-space assignment in buildings. *Automation in Construction*, *40*, 68–83.

Dzeng, R.-J., Wang, W.-C., Hsiao, F.-Y., & Xie, Y.-Q. (2015). An Activity-Based Simulation Model for Assessing Function Space Assignment for Buildings: A Service Performance Perspective. *Computer-Aided Civil and Infrastructure Engineering*, *30*(12), 935–950.

Ekici, B. B., & Aksoy, U. T. (2009). Prediction of building energy consumption by using artificial neural networks. *Advances in Engineering Software*, *40*(5), 356–362.

Enmetric Systems. (n.d.). Retrieved from https://www.enmetric.com/

Erickson, V. L., Lin, Y., Kamthe, A., Brahme, R., Surana, A., Cerpa, A. E., ... Narayanan, S. (2009). Energy efficient building environment control strategies using real-time occupancy measurements. In *Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings* (pp. 19–24). New York, NY, USA: Association for Computing Machinery.

Ethiraj, S. K., & Garg, P. (2012). The Division of Gains from Complementarities in Human-Capital-Intensive Activity. *Organization Science*, *23*(3), 725–742.

Feng, X., Yan, D., & Hong, T. (2015). Simulation of occupancy in buildings. *Energy and Buildings*, *87*, 348–359.

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*(3), 432–441.

Gächter, S., Starmer, C., & Tufano, F. (2015). Measuring the Closeness of Relationships: A Comprehensive Evaluation of the 'Inclusion of the Other in the Self' Scale. *PloS ONE*, *10*(6), e0129478.

Galasiu, A. D., Newsham, G. R., Suvagau, C., & Sander, D. M. (2007). Energy saving lighting control systems for open-plan offices: A field study. *LEUKOS*, *4*(1), 7-29.

Ghahramani, A., Dutta, K., & Becerik-Gerber, B. (2018). Energy trade off analysis of optimized daily temperature setpoints. *Journal of Building Engineering*, *19*, 584–591.

Gill, Z. M., Tierney, M. J., Pegg, I. M., & Allan, N. (2010). Low-energy dwellings: the contribution of behaviours to actual performance. *Building Research & Information*, *38*(5), 491-508.

Grimard, T. (2016). *Raising the IQ of the Buildings We Already Have.* Retrieved from `https://www.enr.com/articles/39703-raising-the-iq-of-the-buildings-we-already-have`

Guerra Santin, O., Itard, L., & Visscher, H. (2009). The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock. *Energy and Buildings*, *41*(11), 1223–1232.

Gulbinas, R., Khosrowpour, A., & Taylor, J. (2015). Segmentation and Classification of Commercial Building Occupants by Energy-Use Efficiency and Predictability. *IEEE Transactions on Smart Grid*, *6*(3), 1414–1424.

Gulbinas, R., & Taylor, J. E. (2014). Effects of real-time eco-feedback and organizational network dynamics on energy efficient behavior in commercial buildings. *Energy and Buildings*, *84*, 493–500.

Gunay, H. B., Fuller, A., O'Brien, W., & Beausoleil-Morrison, I. (2016). Detecting Occupants' Presence in Office Spaces: A Case Study. In *eSIM Conference Proceedings.*

Gunay, H. B., O'Brien, W., & Beausoleil-Morrison, I. (2013). A critical review of observation studies, modeling, and simulation of adaptive occupant behaviors in offices. *Building and Environment*, *70*, 31–47.

Guo, X., Tiller, D., Henze, G., & Waters, C. (2010). The performance of occupancy-based lighting control systems: A review. *Lighting Research & Technology*, *42*(4), 415–431.

Hallac, D., Park, Y., Boyd, S., & Leskovec, J. (2017). Network Inference via the Time-Varying Graphical Lasso. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 205–213). New York, NY, USA: ACM.

Henry, A. D. (2011). Ideology, power, and the structure of policy networks. *Policy Studies Journal*, *39*(3), 361–383.

Hillier, B., & Hanson, J. (1984). *The social logic of space.* Cambridge University Press. Retrieved from http://www.openbibart.fr/item/display/10068/759771

Hoes, P., Trcka, M., Hensen, J. L., & Bonnema, B. H. (2011). Optimizing building designs using a robustness indicator with respect to user behavior. In *Proceedings of Building Simulation 2011: 12th Conference of International Building Performance Simulation Association* (pp. 1710–1717). Retrieved from http://ibpsa.org/proceedings/BS2011/P_1560.pdf

Hong, T., & Lin, H.-w. (2012). Occupant Behavior: Impact on Energy Use of Private Offices. In *Proceedings of ASIM IBSPA Asia Conference.*

Housman, M., & Minor, D. (2016). *Workplace Design: The Good, the Bad, and the Productive* (Harvard Business School Working Papers No. 16-147). Harvard Business School. Retrieved from https://ideas.repec.org/p/hbs/wpaper/16-147.html

Ingle, A., Moezzi, M., Lutzenhiser, L., & Diamond, R. (2014). Better home energy audit modelling: incorporating inhabitant behaviours. *Building Research & Information*, *42*(4), 409–421.

Jain, R. K., Gulbinas, R., Taylor, J. E., & Culligan, P. J. (2013). Can social influence drive energy savings? Detecting the impact of social influence on the energy consumption behavior of networked users exposed to normative eco-feedback. *Energy and Buildings*, *66*, 119–127.

Jain, R. K., Smith, K. M., Culligan, P. J., & Taylor, J. E. (2014). Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. *Applied Energy*, *123*, 168–178.

Jain, R. K., Taylor, J. E., & Peschiera, G. (2012). Assessing eco-feedback interface usage and design to drive energy efficiency in buildings. *Energy and Buildings*, *48*, 8–17.

Jayarajah, K., Lantra, Z., & Misra, A. (2016). Fusing WiFi and Video Sensing for Accurate Group Detection in Indoor Spaces. In *Proceedings of the 3rd International on Workshop on Physical Analytics - WPA '16* (pp. 49–54). New York, New York, USA: ACM Press.

Jazizadeh, F., Ghahramani, A., Becerik-Gerber, B., Kichkaylo, T., & Orosz, M. (2013). Human-Building Interaction Framework for Personalized Thermal Comfort-Driven Systems in Office Buildings. *Journal of Computing in Civil Engineering*, *28*(1), 2–16.

Jazizadeh, F., Ghahramani, A., Becerik-Gerber, B., Kichkaylo, T., & Orosz, M. (2014). User-led decentralized thermal comfort driven HVAC operations for improved efficiency in office buildings. *Energy and Buildings*, *70*, 398–410.

Jia, M., Srinivasan, R. S., & Raheem, A. A. (2017). From occupancy to occupant behavior: An analytical survey of data acquisition technologies, modeling methodologies and simulation coupling mechanisms for building energy efficiency. *Renewable and Sustainable Energy Reviews*, *68*, 525–540.

Jin, M., Jia, R., Kang, Z., Konstantakopoulos, I. C., & Spanos, C. J. (2014). PresenceSense: zero-training algorithm for individual presence detection based on power monitoring. In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings - BuildSys '14* (pp. 1–10). New York, New York, USA: ACM Press.

Jo, J. H., & Gero, J. S. (1998). Space layout planning using an evolutionary approach. *Artificial Intelligence in Engineering*, *12*, 149–162.

Kabo, F. (2018). The architecture of network collective intelligence: Correlations between social network structure, spatial layout and prestige outcomes in an office. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1753), 20170238.

Kabo, F., Hwang, Y., Levenstein, M., & Owen-Smith, J. (2015). Shared Paths to the Lab: A Sociospatial Network Analysis of Collaboration. *Environment and Behavior*, *47*(1), 57–84.

Kazmi, A. H., O'grady, M. J., Delaney, D. T., Ruzzelli, A. G., & O'hare, G. M. P. (2014). A review of wireless-sensor-network-enabled building energy management systems. *ACM Transactions on Sensor Networks*, *10*(4).

Keeling, T., Clements-Croome, D., Luck, R., & Pointer, P. (2012). How the sensory experience of buildings can contribute to wellbeing and productivity. In *Proceedings of 7th Windsor Conference: The changing context of comfort in an unpredictable world.* Windsor, UK: Network for Comfort and Energy Use in Buildings.

Khan, A., Nicholson, J., Mellor, S., Jackson, D., Ladha, K., Ladha, C., . . . Plötz, T. (2014). Occupancy monitoring using environmental & context sensors and a hierarchical analysis framework. In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings* (pp. 90–99). ACM.

Khashe, S., Heydarian, A., Becerik-Gerber, B., & Wood, W. (2016). Exploring the effectiveness of social messages on promoting energy conservation behavior in buildings. *Building and Environment*, *102*, 83–94.

Khosrowpour, A., Xie, Y., Taylor, J. E., & Hong, Y. (2016). One size does not fit all: Establishing the need for targeted eco-feedback. *Applied Energy*, *184*, 523–530.

Kim, T. W., Rajagopal, R., Fischer, M., & Kam, C. (2013). A knowledge-based framework for automated space-use analysis. *Automation in Construction*, *32*, 165–176.

Kleine-Cosack, C., Hennecke, M. H., Vajda, S., & Fink, G. A. (2010). Exploiting Acoustic Source Localization for Context Classification in Smart Environments. In *International Joint Conference on Ambient Intelligence* (pp. 157–166). Springer, Berlin, Heidelberg.

Klepeis, N. E., Nelson, W. C., Ott, W. R., Robinson, J. P., Tsang, A. M., Switzer, P., ... Engelmann, W. H. (2001). The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. *Journal of Exposure Analysis and Environmental Epidemiology*, *11*(3), 231–252.

Krackhardt, D. (1987). QAP partialling as a test of spuriousness. *Social Networks*, *9*(2), 171–186.

Krackhardt, D., & Hanson, J. R. (1993). Informal networks: The company behind the charts. *Harvard Business Review*, *71*(4), 104 - 111.

Krioukov, A., Dawson-Haggerty, S., Lee, L., Rehmane, O., & Culler, D. (2011). A living laboratory study in personalized automated lighting controls. In *Proceedings of the Third ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings* (pp. 1–6). New York, NY, USA: ACM.

Krukar, J., Dalton, R. C., & Hölscher, C. (2016). Applying HCI Methods and Concepts to Architectural Design (Or Why Architects Could Use HCI Even If They Don't Know It). In N. S. Dalton, H. Schnädelbach, M. Wiberg, & T. Varoudis (Eds.), *Architecture and Interaction: Human Computer Interaction in Space and Place* (pp. 17–35). Cham: Springer International Publishing.

Kwak, J.-y., Varakantham, P., Maheswaran, R., Chang, Y.-H., Tambe, M., Becerik-Gerber, B., & Wood, W. (2014). TESLA: an extended study of an energy-saving agent that leverages schedule flexibility. *Autonomous Agents and Multi-Agent Systems*, *28*(4), 605–636.

Lather, J. I., Logan, T., Renner, K., & Messner, J. I. (2019). Evaluating Generated Layouts in a Healthcare Departmental Adjacency Optimization Problem. In *Computing in Civil Engineering* (pp. 539–546).

Lee, H. Y., Yang, I. T., & Lin, Y. C. (2012). Laying out the occupant flows in public buildings for operating efficiency. *Building and Environment*, *51*, 231–242.

Leicht, E. A., & Newman, M. E. J. (2008). Community structure in directed networks. *Physical Review Letters*, *100*(11).

Leskovec, J., Rajaraman, A., Ullman, J. D., Leskovec, J., Rajaraman, A., & Ullman, J. D. (2014). Dimensionality Reduction. In *Mining of Massive Datasets.*

Li, D. H., Cheung, K., Wong, S., & Lam, T. N. (2010). An analysis of energy-efficient light fittings and lighting controls. *Applied Energy*, *87*(2), 558–567.

Lim, B., van den Briel, M., Thiebaux, S., Backhaus, S., & Bent, R. (2012). HVAC-Aware Occupancy Scheduling. In *AAAI Publications, Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence* (pp. 4249–4250).

Lucon, O., Urge-Vorsatz, D., Zain Ahmed, A., Akbari, H., Bertoldi, P., Cabeza, L., ... Vilarino, M. (2014). *Buildings* (Vol. 33; Tech. Rep.).

Majcen, D., Itard, L., & Visscher, H. (2015). Statistical model of the heating prediction gap in Dutch dwellings: Relative importance of building, household and behavioural characteristics. *Energy and Buildings*, *105*, 43–59.

Majumdar, A., Albonesi, D. H., & Bose, P. (2012). Energy-aware meeting scheduling algorithms for smart buildings. In *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings - BuildSys '12* (p. 161). New York, New York, USA: ACM Press.

Majumdar, A., Zhang, Z., & Albonesi, D. H. (2016). Characterizing the Benefits and Limitations of Smart Building Meeting Room Scheduling. In *2016 ACM/IEEE 7th International Conference on Cyber-Physical Systems (ICCPS)* (pp. 1–10). IEEE.

Malliaros, F. D., & Vazirgiannis, M. (2013). Clustering and community detection in directed networks: A survey. *Physics Reports*, *533*(4), 95 - 142.

Manika, D., Wells, V. K., Gregory-Smith, D., & Gentry, M. (2013). The Impact of Individual Attitudinal and Organisational Variables on Workplace Environmentally Friendly Behaviours. *Journal of Business Ethics*, *126*(4), 1–22.

Marin, A., & Wellman, B. (2016). Social Network Analysis: An Introduction. In J. Scott & P. J. Carrington (Eds.), *The SAGE Handbook of Social Network Analysis* (pp. 11–25). London: SAGE Publications Ltd.

Marsden, P. V. (2012). Recent Developments in Network Measurement. In *Models and Methods in Social Network Analysis.*

Martinaitis, V., Zavadskas, E. K., Motuzien, V., & Vilutien, T. (2015). Importance of occupancy information when simulating energy demand of energy efficient house: A case study. *Energy and Buildings*, *101*, 64–75.

Masoso, O., & Grobler, L. (2010). The dark side of occupants' behaviour on building energy use. *Energy and Buildings*, *42*(2), 173–177.

Meier, A. (2006). Operating buildings during temporary electricity shortages. *Energy and Buildings*, *38*(11), 1296–1301.

Melfi, R., Rosenblum, B., Nordman, B., & Christensen, K. (2011). Measuring building occupancy using existing network infrastructure. In *2011 International Green Computing Conference and Workshops, IGCC 2011.*

Menassa, C. C., Kamat, V., & Lee, S. (2013). A Conceptual Framework to Optimize Building Energy Consumption by Coupling Distributed Energy Simulation and Occupancy Models. *Journal of Computing in Civil Engineering*(608), 130217184323001.

Menezes, A. C., Cripps, A., Bouchlaghem, D., & Buswell, R. (2012). Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap. *Applied Energy*, *97*, 355–364.

Menzel, K., Pesch, D., O'Flynn, B., Keane, M., & O'Mathuna, C. (2008). Towards a Wireless Sensor Platform for Energy Efficient Building Operation. *Tsinghua Science and Technology*, *13*(SUPPL. 1), 381–386.

Meyn, S., Surana, A., Lin, Y., Oggianu, S. M., Narayanan, S., & Frewen, T. A. (2009). A sensor-utility-network method for estimation of occupancy in buildings. In *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference* (pp. 1494–1500). IEEE.

Milenkovic, M., & Amft, O. (2013). An opportunistic activity-sensing approach to save energy in office buildings. In *Proceedings of the Fourth International Conference on Future Energy Systems* (pp. 247–258). New York, NY, USA: Association for Computing Machinery.

Misra, S., Cheng, L., Genevie, J., & Yuan, M. (2016). The iPhone Effect. *Environment and Behavior*, *48*(2), 275–298.

Mourshed, M., Shikder, S., & Price, A. D. (2011). Phi-array: A novel method for fitness visualization and decision making in evolutionary design optimization. *Advanced Engineering Informatics*, *25*(4), 676–687.

Nagarathinam, S., Vasan, A., Sarangan, V., Jayaprakash, R., & Sivasubramaniam, A. (2018). Good set-points make good neighbors: User seating and temperature control in uberized workspaces. In *Proceedings of the 5th Conference on Systems for Built Environments* (pp. 144–147). New York, NY, USA: Association for Computing Machinery.

Nagy, D., Lau, D., Locke, J., Stoddart, J., Villaggi, L., Wang, R., ... Benjamin, D. (2017). Project Discover: An Application of Generative Design for Architectural Space Planning. In *Proceedings of the Symposium on Simulation for Architecture and Urban Design.* San Diego, CA, USA: Society for Computer Simulation International.

Neges, M., Koch, C., König, M., & Abramovici, M. (2017). Combining visual natural markers and IMU for improved AR based indoor navigation. *Advanced Engineering Informatics*, *31*, 18–31.

Nguyen, T. A., & Aiello, M. (2012). Beyond Indoor Presence Monitoring with Simple Sensors. In *Proceedings of the 2nd International Conference on Pervasive Embedded Computing and Communication Systems (PECCS-2012)* (pp. 5–14).

Norford, L., Socolow, R., Hsieh, E., & Spadaro, G. (1994). Two-to-one discrepancy between measured and predicted performance of a low-energy' office building: insights from a reconciliation based on the DOE-2 model. *Energy and Buildings*, *21*(2), 121–131.

O'Brien, W., & Gunay, H. B. (2014). The contextual factors contributing to occupants' adaptive comfort behaviors in offices A review and proposed modeling framework. *Building and Environment*, *77*, 77–87.

Olguin Olguin, D., & Pentland, A. S. (2010). Sensor-based organisational design and engineering. *International Journal of Organisational Design and Engineering*, *1*(1/2), 69.

Pan, W., Dong, W., Cebrian, M., Kim, T., Fowler, J. H., & Pentland, A. S. S. (2012). Modeling Dynamical Influence in Human Interaction: Using data to make better inferences about influence within social systems. *IEEE Signal Processing Magazine*, *29*(2), 77–86.

Park, J. Y., Ouf, M. M., Gunay, B., Peng, Y., O'Brien, W., Kjærgaard, M. B., & Nagy, Z. (2019). A critical review of field implementations of occupant-centric building controls. *Building and Environment*, *165*, 106351.

Parys, W., Saelens, D., & Hens, H. (2009). Impact of Occupant Behaviour on Lighting Energy Use. *Building Simulation*(Dc), 1143–1150.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.

Peponis, J., Bafna, S., Bajaj, R., Bromberg, J., Congdon, C., Rashid, M., ... Zimring, C. (2007). Designing space to support knowledge work. *Environment and Behavior*, *39*(6), 815–840.

Perry-Smith, J. E., & Mannucci, P. V. (2017). From creativity to innovation: The social network drivers of the four phases of the idea journey. *Academy of Management Review*, *42*(1), 53-79.

Petersen, S., & Svendsen, S. (2010). Method and simulation program informed decisions in the early stages of building design. *Energy and Buildings*, *42*(7), 1113–1119.

Poll, S., & Teubert, C. (2012). Pilot Study of a Plug Load Management System: Preparing for Sustainability Base. In *2012 IEEE Green Technologies Conference* (pp. 1–6). IEEE.

Potter, G. E., Smieszek, T., & Sailer, K. (2015). Modeling workplace contact networks: The effects of organizational structure, architecture, and reporting errors on epidemic predictions. *Network Science*, *3*(3), 298–325.

Powell, W. W., & Snellman, K. (2004). The Knowledge Economy. *Annual Review of Sociology*, *30*(1), 199–220.

Raftery, P., Keane, M., & Costa, A. (2011). Calibrating whole building energy models: Detailed case study using hourly measured data. *Energy and Buildings*, *43*(12), 3666–3679.

Razavialavi, S., & Abourizk, S. M. (2017). Site Layout and Construction Plan Optimization Using an Integrated Genetic Algorithm Simulation Framework. *Journal of Computing in Civil Engineering*, *31*, 04017011 (1–10).

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (GDPR). (2016). European Union. Retrieved from http://data.europa.eu/eli/reg/2016/679/2016-05-04

Revenue and expense data, University of California. (n.d.). Retrieved from https://www.universityofcalifornia.edu/infocenter/revenue-and-expense-data

Roetzel, A. (2015). Occupant behaviour simulation for cellular offices in early design stagesArchitectural and modelling considerations. *Building Simulation*, *8*(2), 211–224.

Roetzel, A., Tsangrassoulis, A., & Dietrich, U. (2014). Impact of building design and occupancy on office comfort and energy performance in different climates. *Building and Environment*, *71*, 165–175.

Roisin, B., Bodart, M., Deneyer, A., & D 'herdt, P. (2008). Lighting energy savings in offices using different control systems and their real consumption. *Energy and Buildings*, *40*, 514–523.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., & Sejdinovic, D. (2019). Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, *5*(11), eaau4996.

Sailer, K. (2011). Creativity as social and spatial process. *Facilities*, *29*(1), 6–18.

Sailer, K., & McCulloh, I. (2012). Social networks and spatial configuration—How office layouts drive social interaction. *Social Networks*, *34*(1), 47–58.

Sailer, K., Pomeroy, R., & Haslem, R. (2015). Data-driven design—Using data on human behaviour and spatial configuration to inform better workplace design. *Corporate Real Estate*, *4*(3), 249 – 262.

Sandryhaila, A., & Moura, J. M. (2014a). Big Data Analysis with Signal Processing on Graphs: Representation and processing of massive data sets with irregular structure. *IEEE Signal Processing Magazine*, *31*(5), 80–90.

Sandryhaila, A., & Moura, J. M. F. (2013). Discrete Signal Processing on Graphs. *IEEE Transactions on Signal Processing*, *61*(7), 1644–1656.

Sandryhaila, A., & Moura, J. M. F. (2014b). Discrete signal processing on graphs: Frequency analysis. *IEEE Transactions on Signal Processing*, *62*(12), 3042–3054.

Schoofs, A., Delaney, D. T., O'Hare, G. M. P., & Ruzzelli, A. G. (2011). COPOLAN. In *Proceedings of the Third ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings - BuildSys '11* (p. 25). New York, New York, USA: ACM Press.

Schultz, C., Bhatt, M., & Borrmann, A. (2017). Bridging qualitative spatial constraints and feature-based parametric modelling: Expressing visibility and movement constraints. *Advanced Engineering Informatics*, *31*, 2–17.

Shen, W., Newsham, G., & Gunay, B. (2017). Leveraging existing occupancy-related data for optimal control of commercial office buildings: A review. *Advanced Engineering Informatics*, *33*, 230–242.

Shen, Y., Karimi, K., Law, S., & Zhong, C. (2019). Physical co-presence intensity: Measuring dynamic face-to-face interaction potential in public space using social media check-in records. *PLoS ONE*, *14*(2), e0212004.

Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., & Vandergheynst, P. (2013). The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, *30*(3), 83–98.

Siero, F. W., Bakker, A. B., Dekker, G. B., & Van Den Burg, M. T. (1996). Changing Organizational Energy Consumption Behaviour Through Comparative Feedback. *Journal of Environmental Psychology*, *16*(3), 235–246.

Singhvi, V., Krause, A., Guestrin, C., Garrett, J. H., & Matthews, H. S. (2005). Intelligent light control using sensor networks. In *Proceedings of the 3rd International Conference on Embedded Networked Sensor Systems - SenSys '05* (p. 218). New York, New York, USA: ACM Press.

Smith, V., Sookoor, T., & Whitehouse, K. (2012). Modeling building thermal response to HVAC zoning. *ACM SIGBED Review*, *9*(3), 39–45.

Sonta, A. J., & Jain, R. K. (2020). Optimizing Neighborhood-Scale Walkability. In *Computing in Civil Engineering 2019* (pp. 454–461). Atlanta, GA.

Sonta, A. J., Jain, R. K., Gulbinas, R., Moura, J. M. F., & Taylor, J. E. (2017). $\mathrm{OESP_G}$: Computational Framework for Multidimensional Analysis of Occupant Energy Use Data in Commercial Buildings. *Journal of Computing in Civil Engineering*, *31*(4), 04017017.

Sonta, A. J., Simmons, P. E., & Jain, R. K. (2018). Understanding building occupant activities at scale: An integrated knowledge-based and data-driven approach. *Advanced Engineering Informatics*, *37*, 1–13.

Soriano, D. R., & Huarng, K. H. (2013). Innovation and entrepreneurship in knowledge industries. *Journal of Business Research*, *66*(10), 1964–1969.

Sprott, D. E., Spangenberg, E. R., Block, L. G., Fitzsimons, G. J., Morwitz, V. G., & Williams, P. (2006). The question–behavior effect: What we know and where we go from here. *Social Influence*, *1*(2), 128–137.

Staats, H., Harland, P., & Wilke, H. A. M. (2004). Effecting Durable Change: A Team Approach to Improve Environmental Behavior in the Household. *Environment & Behavior*, *36*(3), 341–367.

Staats, H., van Leeuwen, E., & Wit, A. (2000). A longitudinal study of informational interventions to save energy in an office building. *Journal of Applied Behavior Analysis*, *33*(1), 101–104.

Staddon, S. C., Cycil, C., Goulden, M., Leygue, C., & Spence, A. (2016). Intervening to change behaviour and save energy in the workplace: A systematic review of available evidence. *Energy Research & Social Science*, *17*, 30–51.

Suter, G., Petrushevski, F., & Šipetić, M. (2014). Operations on network-based space layouts for modeling multiple space views of buildings. *Advanced Engineering Informatics*, *28*(4), 395–411.

Taherian, S., Pias, M., Coulouris, G., & Crowcroft, J. (2010). Profiling energy use in households and office spaces. In *Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking* (pp. 21–30). ACM.

Tanikawa, H., & Hashimoto, S. (2009). Urban stock over time: spatial material stock analysis using 4d-GIS. *Building Research & Information*, *37*(5-6), 483–502.

Tichy, N., & Fombrun, C. (1979). Network Analysis in Organizational Settings. *Human Relations*, *32*(11), 923–965.

Tomé, A., Kuipers, M., Pinheiro, T., Nunes, M., & Heitor, T. (2015). Spaceuse analysis through computer vision. *Automation in Construction*, *57*, 80–97.

Trivedi, M., Huang Kohsia, & Mikic, I. (2000). Intelligent environments and active camera networks. In *Smc 2000 Conference Proceedings. 2000 IEEE International Conference on Systems, Man and Cybernetics* (Vol. 2, pp. 804–809).

United Nations. (2014). *World Urbanization Prospects 2014* (Tech. Rep.).

U.S. Department of Energy. (n.d.). *EnergyPlus.*

U.S. Energy Information Administration. (2016a). *April 2016 Monthly Energy Review* (Tech. Rep.). Retrieved from http://www.eia.gov/totalenergy/data/monthly/archive/00351604.pdf

U.S. Energy Information Administration. (2016b). *International Energy Outlook 2016* (Tech. Rep.). EIA. Retrieved from http://www.eia.gov/forecasts/ieo/

Uzzi, B., & Spiro, J. (2005). Collaboration and creativity: The small world problem. *American Journal of Sociology*, *111*(2), 447–504.

Vischer, J. C. (2008). Towards a user-centred theory of the built environment. *Building Research & Information*, *36*(3), 231–240.

Waber, B., Magnolfi, J., & Lindsay, G. (2014). Workspaces that move people. *Harvard Business Review*, *92*(10), 6877, 121.

Wang, Z., Wang, Y., Zeng, R., Srinivasan, R. S., & Ahrentzen, S. (2018). Random forest based hourly building energy prediction. *Energy and Buildings*, *171*, 11–25.

Weng, T., & Agarwal, Y. (2012). From Buildings to Smart BuildingsSensing and Actuation to Improve Energy Efficiency. *IEEE Design & Test of Computers*, *29*(4), 36–44.

Weng, T., Balaji, B., Dutta, S., Gupta, R., & Agarwal, Y. (2011). Managing plug-loads for demand response within buildings. In *Proceedings of the Third ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings - BuildSys '11* (p. 13). New York, New York, USA: ACM Press.

Williams, A., Atkinson Pe, B., Phd, K. G., Pe, E. P., & Rubinstein Fies, F. (2012). Lighting Controls in Commercial Buildings. *LEUKOS*, *8*(3), 161–180.

Wineman, J. D., Kabo, F. W., & Davis, G. F. (2009). Spatial and social networks in organizational innovation. *Environment and Behavior*, *41*(3), 427–442.

Xu, X., Taylor, J. E., & Pisello, A. L. (2014). Network synergy effect: Establishing a synergy between building network and peer network energy conservation effects. *Energy and Buildings*, *68*, 312–320.

Yan, D., O'Brien, W., Hong, T., Feng, X., Gunay, H. B., Tahmasebi, F., & Mahdavi, A. (2015). Occupant behavior modeling for building performance simulation: Current state and future challenges. *Energy and Buildings*, *107*, 264–278.

Yang, Z., Ghahramani, A., & Becerik-Gerber, B. (2016). Building occupancy diversity and HVAC (heating, ventilation, and air conditioning) system energy efficiency. *Energy*, *109*, 641–649.

Yang, Z., Li, N., Becerik-Gerber, B., & Orosz, M. (2012). Proceedings of the 2012 Symposium on Simulation for Architecture and Urban Design. In *Proceedings of the 2012 Symposium on Simulation for Architecture and Urban Design* (p. 2). Orlando, FL: Society for Computer Simulation International.

Yu, W., Li, B., Jia, H., Zhang, M., & Wang, D. (2015). Application of multi-objective genetic algorithm to optimize energy efficiency and thermal comfort in building design. *Energy and Buildings*, *88*, 135–143.

Yuan, Y., Alabdulkareem, A., & Pentland, A. S. (2018). An interpretable approach for social network formation among heterogeneous agents. *Nature Communications*, *9*(1), 4704.

Zhao, J., Lasternas, B., Lam, K. P., Yun, R., & Loftness, V. (2014). Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining. *Energy and Buildings*, *82*, 341–355.