Inferring Occupant Ties

Automated Inference of Occupant Network Structure in Commercial Buildings

Andrew J. Sonta
Stanford University
Stanford, California
asonta@stanford.edu

Rishee K. Jain
Stanford University
Stanford, California
rishee.jain@stanford.edu

ABSTRACT

To design and manage office buildings that are both energy-efficient and productive work environments, we need a better understanding of the relationship between building and occupant systems. Past data-driven building research has focused on energy efficiency and occupant comfort, but little work has used building sensor data to understand occupant organizational behavior and dynamics in buildings. In this initial work, we present a methodology for using distributed plug load energy consumption sensors to infer the social/organizational network of occupants (*i.e.*, the relationships among occupants in a building). We demonstrate how plug load data can be used to model activities, and we introduce how statistical methods—in particular, the graphical lasso and the influence model—can be used to learn network structure from time-series activity data. We apply our method to a seven-person office environment in Northern California, and we compare the inferred networks to ground truth spatial, social, and organizational networks obtained through validated survey questions. In the end, a better understanding of how occupants organize and utilize spaces could enable more contextual control and co-optimization of building-human systems.

CCS CONCEPTS

• **Applied computing** → *Engineering*; *Sociology*; • **Computer systems organization** → *Sensor networks*;

KEYWORDS

Social networks, network inference, organizational theory, building management, energy efficiency

ACM Reference Format:
Andrew J. Sonta and Rishee K. Jain. 2018. Inferring Occupant Ties: Automated Inference of Occupant Network Structure in Commercial Buildings. In *The 5th ACM International Conference on Systems for Built Environments (BuildSys '18), November 7–8, 2018, Shenzen, China.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3276774.3276779

1 INTRODUCTION

Commercial office buildings fundamentally exist to enable effective work. Successful office buildings—through their design and management—accomplish this by enhancing certain qualities of the occupant experience, including comfort and productivity. Given the large environmental impact of buildings, good office buildings should also aim to be as energy-efficient as possible.

The new paradigm of sensor data availability in buildings has given researchers more avenues for understanding the operation of buildings through these lenses of environmental performance and human activity. Recent work at the building-human interface has focused on understanding the energy and comfort implications of building operation, such as occupancy-driven operation of HVAC and lighting (*e.g.*, [7]). Fewer studies have focused on using sensor data to model human activity patterns and the natural structure of occupant relationships in commercial buildings. As researchers in the field of organizational behavior have long noted, understanding these relationships can enable more effective space management, for example by suggesting new office layouts that improve workplace satisfaction [9]. Organizational relationships, or *ties*, are typically modeled through surveys that take considerable time and effort to administer. Often, ties are not measured at all, leaving managers with simple organizational charts that describe workforce breakdowns by department or project and lack any subtle insights into the true nature of office relationships. An understanding of occupant ties has also been shown to be useful for reducing energy consumption in office spaces. For example, the efficacy of information campaigns targeted at reducing energy consumption through individual behavior can be largely attributed to social network structure [1]. In recent work [10], we have found that ties can be used to suggest spatial shifts in occupant layouts to more closely match occupant behavior with energy-intensive building systems, thereby reducing overall energy consumption.

Among organizational leaders, the importance of understanding the structure of organizations is well-known. For the University of California system—whose 2016/17 operating budget is public data—total employee salaries, wages, and benefits were roughly 74 times more expensive than utility bills, underscoring the notion that organizations are rational if they prioritize the productivity of their workforce over energy efficiency. While changes to occupant behavior and layout can be key areas for reducing energy consumption, managers would be unlikely to make changes if they worry about disruptions to productivity. At the same time, new research is showing that changes in spatial configurations of offices can improve employee wellbeing and communication [5, 9]. As a result, suggestions for new layouts could be improved if they are made together with an understanding of their effects on work. Using sensor data to gain insight into the occupant network through automatic inference can enable new methods for co-optimizing building and human systems that are fundamentally intertwined.

BuildSys '18, November 7–8, 2018, Shenzen, China

ACM ISBN 978-1-4503-5951-1/18/11…$15.00
https://doi.org/10.1145/3276774.3276779

2 RELATED WORK

In the building energy domain, recent work has considered the occupant network as it relates to energy-related behavior and decision-making among occupants. [1, 10]. In the domain of workspace and organizational theory, researchers have noted an intimate relationship between office design/layout and occupant satisfaction and performance [5, 9]. In fact, recent work has pointed to the notion that spatial configuration can heavily impact key indicators of productivity, such as collaboration [5]. An accurate picture 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 occupant performance [9].

While this previous work has noted the importance of understanding the occupant network for energy and occupant performance, little work has been focused on inferring the true occupant network. Some statistical and data mining tools have been proposed as methods specifically for inferring network structure from time-series data. These methods have typically been applied to biostatistical problems [2], though some recent work has considered the problem of inferring social networks from time series data about human activity [8]. In this paper, we adopt network inference methods for the problem of inferring the occupant network structure from distributed plug-load energy sensors—sensors which are becoming ubiquitous and, as discussed in our previous work, can be used to model individual activity states at the desk level [11].

3 METHODOLOGY

In this section, we introduce a two-step process for inferring the occupant network from raw sensor data. The first step makes use of a method introduced in [11]: raw sensor data are collected from distributed plug-load power strip sensors, and these data are transformed into abstracted states of occupant activities. In the second step, two different models for measuring occupant activities are adapted from the literature and introduced for estimating the network relationships among occupants as defined by their activities. We note that collecting and analyzing data related to occupant network relationships comes with several risks and ethical concerns if misused (e.g., loss of privacy, potential social embarrassment). In order to minimize such risks and concerns, we collected and analyzed data in accordance with the Institutional Research Board (IRB) and ACM's Code of Ethics which included creating a transparent process for obtaining consent for data collection, minimizing personal information collected and automating anonymization of data during collection and analysis.

3.1 Determining occupant activities through plug load energy data

Consistent with previous work [11], we define a time series of plug load energy use collected at the desk level for each occupant: $\mathbf{X}_{i,d} = \{x_1, ..., x_T\}$ 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 total number of time steps at which data are collected in a single day (*e.g.*, if data are collected at 15-min intervals, $T = 96$). For the full dataset, we complete a component selection process based on variational Bayesian inference to determine the number of activity states present in the plug load data (the method is designed to adapt to different building settings, so that the number of states can be vary among study areas). An activity state can be defined as abstracted and categorized information describing occupant behavior based on plug load energy consumption. Once the number of activity states is inferred, we complete a classification process that ascribes each plug load energy datum to an activity state. The result is an abstracted time series that describes overall changes in activity states for all occupants in the study: $\mathbf{X}_{i,d} \mapsto \mathrm{S}_{i,d}$ where S contains the activity states. At each time interval, all occupants are classified into one of the same number of states. For complete details on this state classification method, we refer the reader to [11]. In this work, we have found that plug load sensors provide accuracy comparable to sensors specifically designed for occupant detection, and we have shown that shifts between states correspond accurately which actual changes in behavior, such as going to a meeting.

3.2 Estimating the occupant network

Given time series data about occupant activities, the next step is to infer the occupant network as defined by relative similarities in the activity data. We define an occupant network as a graph $G = (\mathcal{V}, \mathbf{A})$, where $\mathcal{V}$ is the set of occupants and $\mathbf{A}$ is the adjacency matrix of the graph. We explore two options for inferring the adjacency matrix: the graphical lasso, which estimates the inverse covariance matrix—often interpreted as a graphical adjacency matrix [2]; and the influence model, which estimates an 'influence matrix' that is commonly used to describe tie strengths in a network [8]. In this previous work, each of these methods has been shown to scale well to large networks and large time series.

Graphical lasso: The graphical lasso was developed as a method for inferring sparse undirected graphical models—also known as Markov random fields—through L_1 (lasso) regularization. In the literature [2], the data are defined as N multivariate observations with dimension p, mean μ, and covariance Σ. In our case, $N = D \cdot T$ (the total number of time steps), and $p = I$ (the number of workstations/individuals). The graphical lasso makes use of coordinate descent to estimate the inverse covariance matrix (Σ^{-1}), which is often considered as the adjacency matrix in a Markov random field.

Influence model: The influence model, discussed in [8], models the interaction among entities quite differently. The model is based on a generally coupled Hidden Markov Model (HMM), in which the state of each entity (in our setting, S_i) at any given time point t is determined by the state of all entities $\mathrm{S}_{1,...,I}$ at the previous time point $t-1$. A graphical representation of the Influence Model is shown in Figure 1. The authors use Expectation Maximization to estimate the parameters of the model. One of the key parameters of the model that is learned is the matrix $\mathbf{R}$—the 'influence matrix'—often interpreted as an adjacency matrix in a network.

The output of each model is a matrix that we consider as the adjacency matrix defining a weighted, directed network. For consistency, we refer to these matrices as $\mathbf{A}^{\text{glasso}}$ and $\mathbf{A}^{\text{infl}}$, where the entry $\mathbf{A}_{i,j}$ represents the strength of the tie *from* occupant i *to* occupant j. The two models proposed here for learning network structure are based on fundamentally different assumptions. In the graphical lasso, the entries of the $\mathbf{A}^{\text{glasso}}$ can be interpreted as measures of conditional dependence (*i.e.*, if the entry i, j is zero, entities i and j are conditionally independent given all other variables).

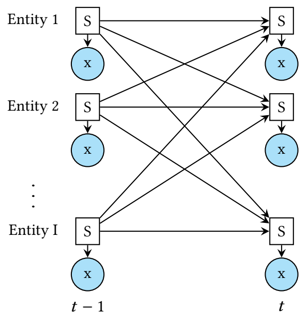

Figure 1: Influence model schematic with I entities (adapted from [8]), where S indicates state, and x indicates the signal.

In the influence model, the entries of $\mathbf{A}^{\text{infl}}$ can be interpreted as the strength of influence between two entities given the time-step-dependent HMM assumptions embedded in the model. Fitting each of these models to the data may provide different insights (*e.g.*, the graphical lasso may effectively model structural relationships, while the influence model may be better at capturing spontaneous trends in behavior, such as when two occupants take a break together).

4 RESULTS AND DISCUSSION

To analyze the performance of our occupant network inference methodology, we applied it to a dataset from a three-room, seven-person office setting in Northern California. We collected plug load energy consumption data at 15-minute intervals through HOBO data loggers for a two-week period at each of the seven occupants' individual workstations. Each workstation included a computer and possibly a monitor and other small office loads (e.g., phone charger). We applied the network inference method from section 3, producing two networks defined by adjacency matrices $\mathbf{A}^{\text{glasso}}$ and $\mathbf{A}^{\text{infl}}$. We also collected ground truth spatial, social, and organizational network data through an online survey (discussed below) to benchmark the inferred network against validated methods for capturing strengths of relationships.

We assume that the ground truth network can be characterized by three equally important relationship components: spatial, social, and organizational. Previous work has suggested these three components are fundamental to the similarities and dissimilarities in occupant behavior [10], but future work should consider how each individual network component relates to the inferred network structure. We refer to each ground truth network component as $\mathbf{A}^{*\text{spatial}}$, $\mathbf{A}^{*\text{social}}$, and $\mathbf{A}^{*\text{organizational}}$.

Spatial dimension: To embed the spatial dimension in the ground truth network, we set the entry $\mathbf{A}^{*\text{spatial}}_{i,j} = \mathbf{A}^{*\text{spatial}}_{j,i} = 1$ if occupants i and j are situated in the same office, and 0 otherwise.

Social dimension: For the social dimension, we used the results from a survey question answered by all occupants. We adopted the 'inclusion of the other in the self' scale [3], which has been shown to be effective in measuring the closeness of social relationships. In the survey, each occupant can choose any value between 1 and 7 to describe their perception of the closeness of their relationship to all other occupants. These values are then linearly scaled between 0 and 1, and they become the entries in the $\mathbf{A}^{*\text{social}}$ matrix, where occupant i's response about occupant j becomes the entry $\mathbf{A}^{*\text{social}}_{i,j}$.

Organizational dimension: We adopt the methods introduced in Krackhardt & Hanson [6] to measure the structure of the organizational network among the seven occupants. The survey asks occupants whom else they (1) share information with, (2) seek technical advice from, and (3) seek personal advice from. The relationship is interpreted to be 0 if none of the three are true and 1 if all three are true, with a linear scale between 0 and 1 if one or two of the characteristics of organizational relationship are true. Occupant i's response about occupant j becomes the entry $\mathbf{A}^{*\text{orgnizational}}_{i,j}$.

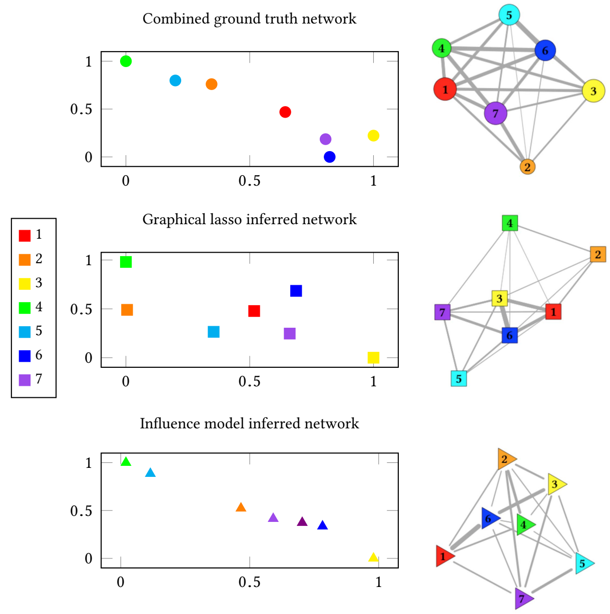

Figure 2: Network comparison using *node2vec* representations in $\mathbb{R}^2$ (the two axes represent the two dimensions). Colors represent the occupant indices.

We combined the individual components of the overall ground truth network using a weighted sum function with the three component networks equally weighted, and we refer to this combined overall ground truth network as $\mathbf{A}^*$. It is difficult to directly compare the overall structure of two networks, though certain network-level techniques (*e.g.*, community detection) or node-level techniques (*e.g.*, degree centrality) can be used. However, we choose to compare the inferred adjacency matrices to the ground truth network by adopting the *node2vec* algorithm introduced in [4]. The purpose of the *node2vec* algorithm is to map nodes from a network to n-dimensional feature representations based on the topology of the network, where n is chosen by the user. The algorithm can be used for dimensionality reduction and enables visual and computational comparisons between networks, where: the smaller the Euclidean distances between two nodes in the feature space in $\mathbb{R}^n$, the more similar the two nodes are. The results of applying the *node2vec* algorithm with $n = 2$ (for ease of plotting) to each of the three networks ($\mathbf{A}^*$, $\mathbf{A}^{\text{glasso}}$, and $\mathbf{A}^{\text{infl}}$) are shown in Figure 2. The overall trends in feature representations for each of the networks are quite similar in Figure 2. Occupants 1 and 7 tend to be centrally located in all 2-dimensional representations. If we define a simple centrality measure for an occupant in this $\mathbb{R}^2$ space as the sum of Euclidean distances between that occupant and all others, in all three networks occupant 1 has the smallest value and occupant 7 has either the second or third smallest value. Knowing the real context of the office environment, this trend makes intuitive sense: occupant 1 is the director of the group, and occupant 7 is the highest-ranking member in the organizational structure, who is often relied on for work-related information and advice. Occupant 6, however, is also centrally located near occupants 1 and 7 in the 2-dimensional representation of the nodes. While this occupant has no structural centrality in the organizational network, inspection of the overall ground truth network reveals that this occupant has relatively strong ties as measured by degree centrality in the ground truth network. Occupant 6's in-degree centrality ($\sum_i \mathbf{A}^*_{i,6}$) is larger than the mean in-degree centrality, and their out-degree centrality ($\sum_i \mathbf{A}^*_{6,i}$) is the largest. The survey data shows that this occupant's degree centrality results in large part from the social and spatial components of the ground truth.

This inclusion of occupant 6 as one of the central nodes in the 2-dimensional representations of the network demonstrates the value of the automated network inference method proposed in this paper. While a manager of an organization might guess that occupant 1 (the director) or 7 (the highest-ranking member) would be central nodes in the network, he or she might not be able to guess that occupant 6 also has a relatively central role. When scaled to large networks, analyses such as these could provide subtle insights into the true structure of the occupant network that could not easily be obtained simply by knowing the structure of the organization. We can also observe from Figure 2 the difference in the vector representations of occupants 2 (orange) and 5 (cyan). While occupant 2 is far away from the central cluster in the graphical lasso network (squares), he or she is more centrally located in the influence model network (triangles). This may suggest that the two models are biased toward capturing different types of relationships given their different assumptions. According to the social component of the ground truth survey, occupants 4 and 5 are close friends, while 2 and 5 are not. Because the influence model assumes that each entity's state at time t is influenced by all entities' states at time $t-1$, it may be more capable of capturing spontaneous similarities in behavior among occupants. For example, given occupant 4 and occupant 5's social relationship, they may be more likely to take breaks or eat lunch together, resulting in the similarities shown in the $\mathbb{R}^2$ representation of $\mathbf{A}^{\text{infl}}$.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we introduce a method for automatically inferring the structure of the *occupant network* from plug load energy consumption data collected at the desk level in an office setting. We have shown that the method is capable of capturing network relationships and centralities that are both expected based on high-level organizational structure and that are 'hidden' in more subtle aspects of occupant relationships (*e.g.*, friendships and spatial configuration). In a 7-person office study, we show that both the graphical lasso and the influence model are capable of capturing these key centralities, and we discuss how some key differences in assumptions might affect each model's network inference output. While the small size of the study limits generalizability of our inferences, our results do demonstrate the potential for this method to effectively and automatically infer organizational structure. Future work is required to methodically demonstrate that inferences like the ones we describe above can be made in larger, more complex settings.

In future work, we aim to more fully explore the difference between each of the modeling techniques as they related to inferred network structure. Additionally, we aim to explore how using data from different times of the day might impact the inferred network structures (*e.g.*, employees might tend to have similar behavioral patterns as their friends during lunchtime, versus similar patterns as their teammates during other hours). Similarly, by extending these inference models to allow the networks to change over time, future work can begin to understand the co-evolution of space design and organizational science and how such insights could inform control paradigms for commercial buildings that co-optimize building and occupant systems. In the end, a deeper understanding of the occupant dynamics within a building could enable the design and management of productive *and* energy-efficient spaces.

REFERENCES

[1] Kyle Anderson, SangHyun Lee, and Carol Menassa. 2012. Effect of Social Network Type on Building Occupant Energy Use. In *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings (BuildSys '12)*. ACM, New York, NY, USA, 17–24. https://doi.org/10.1145/2422531.2422535

[2] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. 2008. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9, 3 (2008), 432–441. https://doi.org/10.1093/biostatistics/kxm045 arXiv:0708.3517

[3] Simon Gächter, Chris Starmer, and Fabio Tufano. 2015. Measuring the Closeness of Relationships: A Comprehensive Evaluation of the 'Inclusion of the Other in the Self' Scale. *PloS one* 10, 6 (jun 2015), e0129478. https://doi.org/10.1371/journal.pone.0129478

[4] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 855–864.

[5] Felichism Kabo, Yongha Hwang, Margaret Levenstein, and Jason Owen-Smith. 2015. Shared Paths to the Lab: A Sociospatial Network Analysis of Collaboration. *Environment and Behavior* 47, 1 (jan 2015), 57–84. https://doi.org/10.1177/0013916513493909

[6] David Krackhardt and Jeffrey R. Hanson. 1993. Informal networks: The company behind the charts. *Harvard Business Review* 71, 4 (jul 1993), 104 – 111.

[7] Andrew Krioukov, Stephen Dawson-Haggerty, Linda Lee, Omar Rehmane, and David Culler. 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 (BuildSys '11)*. ACM, New York, NY, USA, 1–6. https://doi.org/10.1145/2434020.2434022

[8] Wei Pan, Wen Dong, Manual Cebrian, Taemie Kim, James H. Fowler, and Alex S (Sandy) Pentland. 2012. Modeling Dynamical Influence in Human Interaction: Using data to make better inferences about influence within social systems. *IEEE Signal Processing Magazine* 29, 2 (mar 2012), 77–86. https://doi.org/10.1109/MSP.2011.942737

[9] Kerstin Sailer, Ros Pomeroy, and Rosie Haslem. 2015. Data-driven design—Using data on human behaviour and spatial configuration to inform better workplace design. *Corporate Real Estate* 4, 3 (2015), 249 – 262.

[10] Andrew J. Sonta, Rishee K. Jain, Rimas Gulbinas, José M. F. Moura, and John E. Taylor. 2017. OESPG: Computational Framework for Multidimensional Analysis of Occupant Energy Use Data in Commercial Buildings. *Journal of Computing in Civil Engineering* 31, 4 (jul 2017), 04017017. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000663

[11] Andrew J. Sonta, Perry E. Simmons, and Rishee K. Jain. 2018. Understanding building occupant activities at scale: An integrated knowledge-based and data-driven approach. *Advanced Engineering Informatics* 37 (2018), 1–13. https://doi.org/10.1016/j.aei.2018.04.009