

RESEARCH ARTICLE

Role detection in bicycle-sharing networks using multilayer stochastic block models

Jane Carlen^{†,*1}, Jaume de Dios Pont^{†,2}, Cassidy Mentus^{†,2}, Shyr-Shea Chang², Stephanie Wang³ and Mason A. Porter^{2,4*} 

¹Department of Statistics, University of California, Los Angeles, USA, ²Department of Mathematics, University of California, Los Angeles, USA, ³Department of Computer Science and Engineering, University of California, San Diego, USA, and ⁴Santa Fe Institute, USA

*Corresponding author. E-mail: mason@math.ucla.edu

Action Editor: Ulrik Brandes

Abstract

In urban systems, there is an interdependency between neighborhood roles and transportation patterns between neighborhoods. In this paper, we classify docking stations in bicycle-sharing networks to gain insight into the human mobility patterns of three major cities in the United States. We propose novel time-dependent stochastic block models, with degree-heterogeneous blocks and either mixed or discrete block membership, which classify nodes based on their time-dependent activity patterns. We apply these models to (1) detect the roles of bicycle-sharing stations and (2) describe the traffic within and between blocks of stations over the course of a day. Our models successfully uncover work blocks, home blocks, and other blocks; they also reveal activity patterns that are specific to each city. Our work gives insights for the design and maintenance of bicycle-sharing systems, and it contributes new methodology for community detection in temporal and multilayer networks with heterogeneous degrees.

Keywords: stochastic block models, bicycle-sharing networks, multilayer networks, time-dependent networks, transportation systems

1. Introduction

Transportation systems and commuting patterns shape and reveal functional regions in a city (Barbosa *et al.*, 2018; Barthelemy, 2019). The continuing development of network-analysis methods and the increasing availability of transportation data (from a multitude of transportation modes) give exciting opportunities to improve understanding of urban dynamics at both large and small scales. One prominent type of transportation service is bicycle-sharing systems, which can adapt quickly to the needs of travelers. The number of bicycle-sharing programs worldwide has grown rapidly, from 5 in 2005 to 1571 in 2018 (Schmidt, 2018). Over 50 systems were launched in the United States alone between 2010 and 2016, and over 20 bicycle-sharing systems have been launched in France since 2005 (Etienne and Latifa, 2014; Bike share in the U.S, 2018). Many existing bicycle-sharing systems are also growing. For example, the number of stations in New York City's "Citi Bike" system has more than doubled since it debuted in 2013.

[†]Jane Carlen, Jaume de Dios Pont, and Cassidy Mentus are joint lead authors of the paper.

^{*}Now at the *Los Angeles Times*.

Docked bicycle-sharing systems follow a general structure: Groups of bicycles are parked at “stations” (which are also called “docks” or “hubs”) throughout a coverage area, and users withdraw and return bicycles to these stations on demand, with a cost that depends on usage time. A growing number of bicycle-sharing systems are dockless (as are the increasingly prominent e-scooters), so users can park bicycles at any location in a coverage area. In the present paper, we analyze docked systems, but we consider how to adapt our models to dockless systems in Section 7.

Data from bicycle-sharing systems are available from many cities throughout the world (Austwick *et al.*, 2013; Romanillos and Zaltz Austwick, 2016; Wergin and Buehler, 2017; Munoz-Mendez *et al.*, 2018; Romanillos *et al.*, 2018). Such data, which include detailed temporal records and GPS-tracked routes in some cases, are helpful for capturing commuting behavior (Fishman *et al.*, 2013). Analyzing bicycle-sharing data can increase our understanding of urban flows and human commuting. Bicycle-sharing is often used for “last-mile” transportation, bridging the gap between public transportation and a final destination (Griffin and Sener, 2016), and insights into the dynamics of bicycle-sharing systems can help transit systems evolve to meet the needs of changing cities (Ashqar, 2019; Natera Orozco, 2020). Studying bicycle-sharing systems is also helpful for maintaining and expanding them (Ashqar, 2021). For example, it can yield helpful ideas for the efficient redistribution of bicycles, a problem that has received much attention (Shu *et al.*, 2013; Pfrommer *et al.*, 2014; Forma *et al.*, 2015; Singhvi *et al.*, 2015).

In the present paper, we propose two models of time-dependent networks to capture the functional roles of bicycle-sharing stations. We use the lens of mesoscale-structure detection in time-dependent networks (Holme and Saramäki, 2012; Holme, 2015; Fortunato and Hric, 2016). We examine trip histories from bicycle-sharing systems in the form of multilayer networks (Kivelä *et al.*, 2014; Aleta and Moreno, 2019; Porter, 2018) in which each layer is a network of trips that start in a given hour.

We aim to partition each of our networks based on a relational equivalence of nodes (a perspective with a rich history in the social-networks literature (Lorrain and White, 1971; Rossi and Ahmed, 2015)), rather than on high internal traffic within sets of nodes (Munoz-Mendez *et al.*, 2018). The analysis of time-aggregated data can shed light on “community” membership in the latter sense (Austwick *et al.*, 2013) through a partition of a network into contiguous spatial clusters (Munoz-Mendez *et al.*, 2018). However, such an approach ignores how bicycle-sharing usage relates to travel patterns during a day (Fishman *et al.*, 2013). By contrast, our models are designed to detect functional roles of bicycle-sharing stations based on time-dependent usage patterns.

The models that we introduce in this paper are time-dependent extensions of stochastic block models (SBMs) (Snijders and Nowicki, 1997; Nowicki and Snijders, 2001; Karrer and Newman, 2011; Peixoto, 2019). We include parameters to describe intra-block and inter-block traffic that starts in each hour, but we fix the block assignment of each node over time. That is, we treat a bicycle-sharing network as a temporal multilayer network with fixed node identities across layers (Kivelä *et al.*, 2014; Aleta and Moreno, 2019). We use the terminology “time-dependent” to emphasize the time-dependent nature of our application and the associated interpretations of our results, but our models are relevant for studying any multilayer network with the same set of nodes across layers and no interlayer edges. We also note that the layers do not need to be ordered.

We formulate both mixed-membership and discrete-membership models, where nodes can be members of multiple blocks or exactly one block, respectively. Our models are degree-corrected. In an SBM without degree-correction, the expected weight of an edge is determined by the block memberships of its incident nodes. By contrast, in a degree-corrected SBM, the expected weight of an edge also depends on the activity levels (i.e., the degrees) of its incident nodes. We extend the degree-correction that was introduced by Karrer and Newman (2011) to our multilayer SBMs (see Section 3). Incorporating degree-correction when modeling networks with heterogeneous degrees can increase the performance of a model while only slightly increasing its complexity (Karrer and Newman, 2011).

In Section 6, we illustrate that degree-correction enables us to identify blocks based on their functional roles in a network, rather than based on their levels of activity. This is important for bicycle-sharing networks, as stations have heterogeneous numbers of parking spaces for bicycles and different neighborhoods have different baseline levels of bicycle usage. Our models are applicable to directed networks and are easily extendable to undirected networks. Additionally, although our models are inspired by the analysis of bicycle-sharing systems, they are applicable more generally to multilayer networks in which nodes have heterogeneous degrees and belong to fixed classes.

Community detection in multilayer networks is a very active research area, and it has yielded insights into many problems in diverse disciplines (Cranmer *et al.*, 2015; Valdano *et al.*, 2015; Barbillon *et al.*, 2017; Papadopoulos *et al.*, 2018; Kobayashi *et al.*, 2019; Bazzi *et al.*, 2020). Many algorithms for community detection have been generalized to multilayer networks. See, for example, Mucha *et al.* (2010), Yang *et al.* (2011), Paul and Chen (2016), Stanley *et al.* (2016), Valles-Catala *et al.* (2016), Barbillon *et al.* (2017), Jeub *et al.* (2017), and Zhang *et al.* (2017).

The SBMs that we introduce have parameters for block-to-block activity that are distinct for different departing hours; however, the block-membership (i.e., community-membership) parameters are fixed across all hours. Most related SBMs that have been introduced differ in one or both of these features. Yang *et al.* (2011) introduced a discrete-membership SBM in which the parameters for block-to-block activity are fixed with respect to time but block memberships evolve with time. Xu and Hero (2014) and Matias and Miele (2017) proposed related models (for unweighted and weighted networks, respectively). However, unlike in the model of Yang *et al.*, their models do not fix block-to-block activity parameters over time. Zhang *et al.* (2017) developed a time-dependent block model with degree-correction in which block membership does not change with time and the edge dynamics are described by a continuous-time Markov process, with edges that are added or removed between block pairs at constant rates. Mixed-membership SBMs with time-evolving communities have also been developed (Xing *et al.*, 2010; Ho *et al.*, 2011). See Rossetti and Cazabet (2018) for a survey of community-detection methods for time-dependent networks.

Community detection has been applied to urban bicycle-sharing systems using a variety of approaches (Rosvall *et al.*, 2009; Borgnat *et al.*, 2011; Austwick *et al.*, 2013; Akbarzadeh *et al.*, 2018; Munoz-Mendez *et al.*, 2018; Xie and Wang, 2018; He *et al.*, 2019; Kobayashi *et al.*, 2019). Zhu *et al.* (2018) applied *k*-means clustering to undirected, time-dependent usage data from bicycle-sharing systems and other urban systems in New York City. Austwick *et al.* (2013) examined modularity optimization with a directed and spatially-corrected null model to identify communities of bicycle-sharing stations in several cities. However, they detected communities in time-aggregated networks; as they pointed out, there are significant limitations in bicycle-sharing applications when examining community structure while ignoring time-dependent usage patterns. Munoz-Mendez *et al.* (2018) identified communities in hourly slices of bicycle-sharing data from London using an INFOMAP algorithm (Rosvall *et al.*, 2009) that respects the directed nature of edges in the underlying bicycle-trip networks. The changes with time that they discovered in those communities highlight the importance of time of day in the usage patterns of bicycle-sharing systems.

There is some recent work that is closely related to ours. Matias *et al.* (2018) constructed a time-dependent, discrete-membership SBM with fixed blocks over time and applied it to bicycle-sharing networks in London. They detected some functional blocks, but their approach does not incorporate degree-correction. Xie and Wang (2018) employed an approach that does not use an SBM directly, but they were able to successfully partition a bicycle-sharing network to find home and work roles of bicycle-sharing stations. They used the ratio of in-degree to out-degree at different times to discover home–work splits during peak commute times. In the present paper, we observe such splits using our two-block models. A similarity of their approach to ours is that it corrects for degree; a key difference is that they relied on human supervision to determine peak hours, whereas our models implicitly increase the weights of more-active hours in our likelihood function. Etienne and Latifa (2014) used a Poisson mixture model to cluster bicycle-sharing stations in

Paris based on their time-dependent usage profiles. They were able to capture the time-dependent activity patterns of each group of stations, distinguish between arrival and departure activity, control for the overall activity level of a given station (via degree-correction), and associate the groups that they identified with their roles in the city. A major difference between their approach and ours is that we distinguish activity between blocks, which allows us to detect behavior like last-mile commuting that occurs within blocks. We also consider mixed membership, which they proposed as a subject for future work.

Our paper proceeds as follows. In Section 2, we list our data sources and present basic statistical analysis of the data. In Section 3, we introduce our time-dependent SBMs—a mixed-membership model and a discrete-membership model—and we show that they are equivalent up to a constraint. In Section 4, we describe estimation algorithms for our SBMs. In Section 5, we generate synthetic networks using both of our models; we evaluate the fits of our models and a model by Matias *et al.* (2018) to the generated data. In Section 6, we present the results of fitting our models to bicycle-sharing networks in Los Angeles, San Francisco, and New York City. We summarize our results, discuss their implications for bicycle-sharing systems, and suggest areas of further study in Section 7. We show some additional details of our work in the Appendix. We provide code and data to implement our models and replicate the results in our paper at https://github.com/jcarlen/tdsbm_supplementary_material.

2. Data

We examine bicycle-sharing systems in Los Angeles, the Bay Area, and New York City. For Los Angeles, we study only the system's downtown region, which is self-contained in the sense that very few trips connect it to stations that are not downtown. Similarly, for the Bay Area, we consider only the San Francisco network and exclude stations in Mountain View, Palo Alto, and San Jose. (If we applied our models to these nearly disconnected portions of the Bay Area bicycle-sharing network, it would be better to treat them as separate networks than to model them jointly.) Each of our three focal systems covers a dense urban area, but they vary in their sizes and daily usage. Because of this variability and how the data were reported, we study different time periods for each system. We also select our time periods to exclude days that are likely to be extremely hot or extremely cold. Each of the bicycle-sharing systems that we study has an open-data portal, from which we downloaded our data. After cleaning (see our discussion in the next paragraph), our data consist of the following:

- Downtown Los Angeles: October–December 2016; there are 61 stations and 40,130 trips, of which 73.4% are during weekdays (NYCBS, 2017).
- San Francisco: September 2015–August 2016; there are 35 stations and 267,412 trips, of which 92.1% are during weekdays (Bay Area Bike Share, 2017).
- New York City: October 2016; there are 601 stations and 1,551,692 trips, of which 75.6% are during weekdays (LA Metro, 2017).

A trip consists of a user checking out a bicycle from a fixed location (a station that includes multiple parking spaces) and returning it to a station. The data for each trip include the starting time; the ending time; and the starting and ending locations by station ID, latitude, and longitude. Each data set also has a few additional fields about the users; these details include whether or not they have memberships in the bicycle-sharing system, but we do not use this information in our investigation. We clean the data by removing anomalous trips, including extremely short and extremely long trips,¹ and trips that were for testing or maintenance (as indicated in the data). We also exclude a very small number of stations (two in Los Angeles and six in New York City) that did not have at least one departure and at least one arrival during the examined time period.

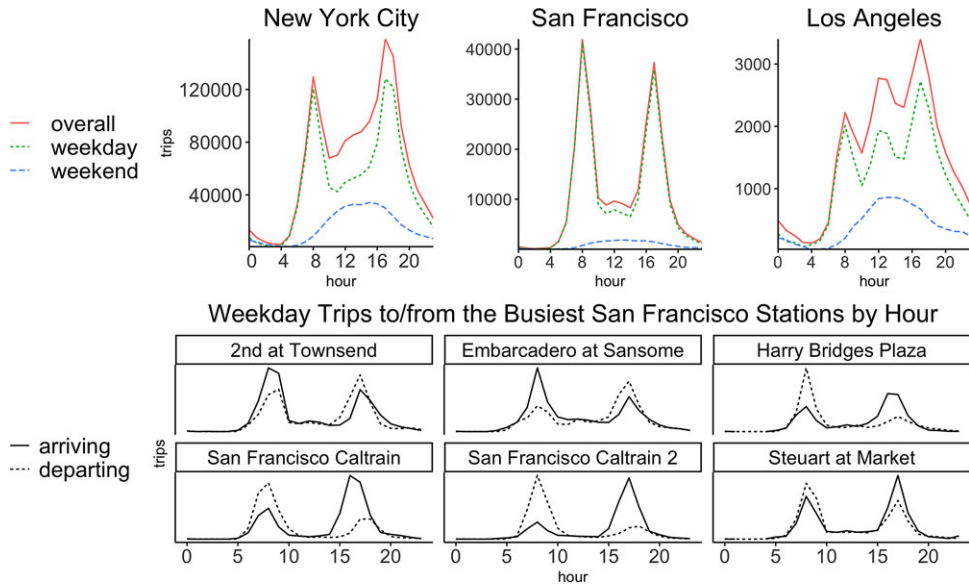


Figure 1. (Top) Total bicycle trips by departing hour for weekdays, weekends, and overall in New York City, San Francisco, and Los Angeles. Hour 0 designates midnight. (Bottom) Trips that arrive (solid curves) and depart (dashed curves) by hour in the six San Francisco stations with the most trips.

Finally, we exclude one station in New York City that was accessible to other stations only by ferry; it was involved in only nine trips during the examined time period. In total, cleaning removes 7.1% of the trips in Los Angeles, 4.5% of the trips in San Francisco, and 1.4% of the trips in New York City.

We retain self-edges, which represent trips that start and end at the same station. Although it is common to remove self-edges when analyzing networks (Newman, 2018), we expect the self-edges in our data to have a very similar data-generating mechanism as trips to nearby stations. As in Karrer and Newman (2011), including self-edges also simplifies some aspects of parameter estimation. When fitting our models, we include only weekday trips, as we observe that weekday and weekend activity follow distinct patterns (see Figure 1). Moreover, weekend activity does not reflect daily commuting behavior. From the data sets, we construct multilayer networks that are both weighted and directed. In our networks, nodes represent stations, edge values encode the number of directed trips from one station to another that begin in a specified hour, and each of the 24 layers encompasses the trips that start in a certain hour.

2.1 Preliminary data analysis

When examining the bicycle-sharing data sets, we observe strong daily usage patterns, which motivate the development of our time-dependent SBMs. In Figure 1, we illustrate these usage patterns by plotting the number of trips by starting hour for each city. In each city, weekday activity spikes during morning and evening commuting hours (and Los Angeles also has a mid-day spike), whereas weekend trips peak in early afternoon. Similar patterns have been observed previously for bicycle-sharing systems in New York City and many other cities (Austwick *et al.*, 2013; Etienne and Latifa, 2014; Taiyeb, 2014; Xie and Wang, 2018; Zhu *et al.*, 2018).

For individual stations, the morning and evening peaks for arriving trips (i.e., in-edges) and departing trips (i.e., out-edges) are often unbalanced. One direction has a stronger morning peak, and the opposite direction has a stronger evening peak. In the bottom panel of Figure 1, we show

examples of this imbalance for the six busiest stations in San Francisco. (In Appendix A.1, we further explore this issue using singular-value decompositions of the trip data.) These patterns suggest how stations are used in commuting, and they motivate our time-dependent identification of stations into “home” and “work” types.

Another characteristic of our data that we incorporate into the design of our models is the strong positive Pearson correlation coefficient between the total (summed over all hours) in-degree and out-degree of each station; it is 0.99 in New York City, 0.98 in San Francisco, and 0.91 in Los Angeles. Accordingly, we apply degree-correction to the overall degree of each station, rather than to the in-degrees and out-degrees separately (see Section 3.1). This correlation is an intrinsic feature of docked bicycle-sharing systems; at some point, bicycles must be returned to a station before any new trips can originate from it. (The use of trucks to redistribute bicycles within a system can loosen this requirement.) More generally, the strong correlation also relates to a famous axiom of human mobility: for each current of travel, there is an associated countercurrent (Ravenstein, 1885).

3. Our stochastic block models

In this section, we introduce our time-dependent mixed-membership stochastic block model (TDMM-SBM) and time-dependent discrete stochastic block model (TDD-SBM).

SBMs are a popular class of statistical network models (Peixoto, 2019). The motivating principle of SBMs is a notion of *stochastic equivalence* in which edges whose incident nodes have the same block membership are identically distributed. It is a standard assumption of SBMs that edge weights are independent, given the block memberships of their associated nodes. More formally, an unweighted (i.e., Bernoulli) random graph Y (with adjacency matrix A) from an SBM with K blocks is defined by

$$P(A_{ij} = 1|G) = \eta_{g_i g_j}, \quad (1)$$

where G (with entries $g_i \in \{1, \dots, K\}$, with $i \in \{1, \dots, N\}$) is a vector of block assignments for the nodes of Y and η is a $K \times K$ matrix of block-to-block edge probabilities. This definition allows directed graphs, in which η can be asymmetric. For early presentations of SBMs, see Holland (1983), Faust and Wasserman *et al.* (1992), Snijders and Nowicki (1997), and Nowicki and Snijders (2001). More recent advances have added flexibility to SBMs. Examples include mixed-membership and overlapping-membership SBMs (Airoldi *et al.*, 2008; Latouche *et al.*, 2011), models with covariates (Leger *et al.*, 2020), degree-corrected SBMs (Karrer and Newman, 2011), and Bayesian implementations (Peixoto, 2019). Applications of SBMs to time-dependent networks include discrete-membership (Yang *et al.*, 2011; Zhang *et al.*, 2017) and mixed-membership (Xing *et al.*, 2010; Ho *et al.*, 2011) models in which the block assignments of nodes are time-dependent and discrete-membership models (without degree-correction) with fixed blocks over time (Xu and Hero, 2014; Matias *et al.*, 2018).

3.1 Time-dependent mixed-membership stochastic block model (TDMM-SBM)

We now describe the framework for our mixed-membership SBM. Let $i, j \in \mathcal{N}$ (with $|\mathcal{N}| = N$) be nodes, which represent bicycle-sharing stations; let $g, h \in \mathcal{K}$ (with $|\mathcal{K}| = K$) be blocks. We treat both the number of nodes and the number of blocks as fixed and given. To each bicycle-sharing network, we associate a three-dimensional array of size (N, N, T) , where T is the number of time slices (i.e., time layers). Our model is also valid when layers have meanings other than time slices. However, because of our application to bicycle-sharing systems, we describe the layers as time layers for convenience. We consider hourly groupings of bicycle trips based on their starting times, and we do not include interlayer edges.² The array entry A_{ijt} is the observed number of trips from station i to station j with a starting time that is at least t and less than $t + 1$. Our networks

are directed, so we count each trip that starts at node i in hour t and ends at node i (i.e., self-edges) exactly once in A_{iit} . We use $\tilde{A}_{ij} = \sum_{t=0}^{23} A_{ijt}$ to denote the edge weights of the associated time-aggregated matrix.

For each node i , there is a latent length- K vector of real numbers $C_{ig} \in [0, 1]$. These vectors, which we will estimate, signify the mixed-membership block assignments of the nodes. The block-assignment parameter C_{ig} indicates the “strength” of node i in block g . For each ordered pair g, h of blocks and each time $t \in \{0, 1, \dots, 23\}$ (where $t = 0$ represents the hour that starts at midnight), there is a parameter ω_{ght} , which we call the “block-connectivity” parameter (i.e., the “block-to-block” activity parameter), that represents the directed activity from block g to block h in layer t . Note that ω_{ght} need not be equal to ω_{hgt} for a directed network; in our application, this captures asymmetries in the numbers of trips with respect to reversing origins and destinations. We also define the notation $\tilde{\omega}_{gh} = \sum_{t=0}^{23} \omega_{ght}$ for our time-aggregated matrix.

For each pair of nodes, i and j , we assume that the number of trips that depart from i in hour t and arrive at j at any time is Poisson-distributed with mean $\mu_{ijt} = \sum_{g,h} C_{ig} \omega_{ght} C_{jh}$. Our use of the Poisson distribution follows Karrer and Newman (2011) and Peixoto (2019), facilitates computation, and is standard for modeling count data (although overdispersion is a concern).

For identifiability (i.e., to allow unambiguous inference), we apply the constraint $\sum_i C_{ig} = 1$ for all g . This does not constrain the set of possible models in terms of realizable mean edge activities μ_{ijt} . Consider a model with unconstrained parameters ω_{ght} and C_{ig} . A model with parameters ω'_{ght} and C'_{ig} such that $C'_{ig} = \frac{C_{ig}}{\sum_j C_{jg}}$ and $\omega'_{ght} = \omega_{ght} \left(\sum_j C_{jg} \right) \left(\sum_j C_{jh} \right)$ gives an equivalent parameterization, because the means of the distributions of edge weights are equal to the those of the model with unconstrained parameters. That is, $\mu'_{ijt} = \sum_{g,h} C'_{ig} \omega'_{ght} C'_{jh} = \sum_{g,h} C_{ig} \omega_{ght} C_{jh} = \mu_{ijt}$.

Given that $\sum_i C_{ig} = 1$, we can think of C_{ig} as the proportion of the total activity of block g from the activity of node i . We interpret C_{ig} relative to C_{ih} as how strongly block g is associated with node i relative to how strongly block h is associated with node i . The expected total number of trips at node i is $\sum_g C_{ig} \sum_{h,t} (\omega_{ght} + \omega_{hgt})$. By contrast, $\sum_g C_{ig}$ is a measure of the activity of node i in which we do not weight C_{ig} by the total activity of its corresponding block. We use $\sum_g C_{ig}$ to determine the size of node i when plotting the block assignments that we infer using the TDMM-SBM (see the left panels of Figures 3 and 6) because it helps ensure that we do not overlook blocks with important usage patterns but relatively low activity. The parameters C_{ig} , with i fixed and $g \in \mathcal{K}$, extend the degree-correction parameter for SBMs that was introduced by Karrer and Newman (2011) to mixed block membership. We elaborate on this connection in Section 3.2, where we introduce a model that specifies that nodes are assigned to only one block.

We now compute the likelihood function that we will optimize to obtain maximum-likelihood estimates (MLEs) of parameters. We assume conditional independence between the Poisson-distributed hourly numbers of trips along each edge, given model parameters, so the likelihood of the data is

$$L(G; \omega, \mathbf{C}) = \prod_{t=0}^{23} \prod_{i,j \in \mathcal{N}} \frac{(\mu_{ijt})^{A_{ijt}}}{A_{ijt}!} \exp(-\mu_{ijt}), \quad (2)$$

where ω and \mathbf{C} are the model parameters (i.e., $\omega = \{\omega_{ght}\}$ and $\mathbf{C} = \{C_{ig}\}$). Note that $\mu_{ijt} = \sum_{g,h} C_{ig} \omega_{ght} C_{jh}$ is a function of these parameters, the set \mathcal{N} of nodes of the network is fixed and given, and the number K of blocks is also fixed and given. The unnormalized log-likelihood is

$$\ell(G; \omega, \mathbf{C}) = \sum_{t=0}^{23} \sum_{i,j \in \mathcal{N}} [A_{ijt} \log(\mu_{ijt}) - \mu_{ijt}], \quad (3)$$

where we omit the constant $-\sum_{i,j,t} \log(A_{ijt}!)$ because it does not affect the maximum of the function. In (3) and throughout this paper, we use “log” to denote the natural logarithm.

3.2 Time-dependent discrete stochastic block model (TDD-SBM)

We derive a single-membership SBM from our mixed-membership SBM by making the extra assumption, for each node $i \in \mathcal{N}$, that $C_{ig} > 0$ for only one block $g \in \mathcal{K}$. For our single-membership SBM, we introduce some new notation to aid our description and promote consistency with the notation in other work (Karrer and Newman, 2011; Zhu et al., 2013). For a given node i , the block g for which $C_{ig} > 0$ is the block g_i that is associated with node i . Therefore, we describe this model as a discrete-membership model. We introduce a scalar parameter $\theta_i = C_{ig_i}$ for each node i ; it is a multilayer extension of the degree-correction parameter in Karrer and Newman (2011). The expectation of the Poisson distribution of the value of an edge from node i to node j at time t is $\theta_i \theta_j \omega_{g_i g_j t} = C_{ig_i} \omega_{g_i g_j t} C_{jg_j} = \mu_{ij t}$. We retain the sum constraints of our mixed-membership model, so $\sum_{i \in g} \theta_i = 1$ for all g .

We compute optimal values of the parameters ω and $\theta = \{\theta_i\}_{i \in \mathcal{N}}$. As in our TDMM-SBM, we take \mathcal{N} and \mathcal{K} to be fixed and given. Again dropping the constant term $-\sum_{i,j,t} \log(A_{ijt}!)$, the log-likelihood of our discrete-membership SBM is

$$\ell(G; \omega, \theta) = \sum_t \sum_{g,h} \sum_{i \in g, j \in h} [A_{ijt} \log \theta_i + A_{jit} \log \theta_j + A_{ijt} \log \omega_{ght} - \theta_i \theta_j \omega_{ght}]. \tag{4}$$

This resembles the degree-corrected SBM in equation (14) of Karrer and Newman (2011), but we have adapted it to directed networks and summed over time layers.

We find explicit formulas for the conditional MLEs of θ_i and ω_{ght} , given the block memberships $\{g_i\}_{i \in \mathcal{N}}$. In the following calculations, the absence of t in the subscript of a parameter and the presence of a tilde designates a sum over all t . Specifically, we define $\tilde{A}_{ij} = \sum_{t=0}^{23} A_{ijt}$ and $\tilde{\omega}_{gh} = \sum_{t=0}^{23} \omega_{ght}$. We differentiate ℓ with respect to ω_{ght} to obtain

$$\frac{\partial}{\partial \omega_{ght}} \ell = \frac{\sum_{i \in g, j \in h} A_{ijt}}{\omega_{ght}} - 1,$$

where we have used the blockwise sum constraints on θ to simplify $\sum_{i \in g, j \in h} \theta_i \theta_j$ to 1. Therefore, the conditional MLE for ω_{ght} is

$$\hat{\omega}_{ght} = m_{ght}, \tag{5}$$

where m_{ght} is the sum of the weights of the edges from nodes in block g to nodes in block h in layer t . That is, $m_{ght} = \sum_{i \in g, j \in h} A_{ijt}$.

We then differentiate ℓ with respect to θ_i to obtain

$$\frac{\partial}{\partial \theta_i} \ell = \frac{\sum_j \tilde{A}_{ij} + \sum_j \tilde{A}_{ji}}{\theta_i} - \sum_h \tilde{\omega}_{g_i h} - \sum_h \tilde{\omega}_{hg_i}.$$

At $\hat{\omega}_{ght}$, the conditional MLE for θ_i is

$$\hat{\theta}_i = \frac{\sum_j (\tilde{A}_{ij} + \tilde{A}_{ji})}{\sum_g (\tilde{m}_{gg_i} + \tilde{m}_{g_i g})} = \frac{k_i}{\kappa_{g_i}}, \tag{6}$$

where $k_i = \sum_j (\tilde{A}_{ij} + \tilde{A}_{ji})$ is the degree of node i (i.e., the sum of the in-degree and the out-degree³ of i) over all time layers, $\tilde{m}_{gh} = \sum_{t=0}^{23} m_{ght}$, and $\kappa_g = \sum_h (\tilde{m}_{gh} + \tilde{m}_{hg})$ is the sum of the in-degrees and out-degrees of all nodes in block g over all time layers. The term $2\tilde{m}_{gg}$ in the equation for κ_g

implies that we count each intra-block edge twice: once for departing from g and once for arriving at g . Similarly, k_i includes the term $2\tilde{A}_{ii}$, so we count self-edges twice in this term.

The parameter MLE $\hat{\theta}_i$ in our TDD-SBM for directed, multilayer networks is analogous to the MLE of a degree-correction parameter in a degree-corrected SBM for undirected networks with one layer (Karrer and Newman, 2011). Indeed, the MLE for the degree-correction parameter of a node in the latter model is the degree of that node divided by the sum of the degrees of all nodes in its block. Another similarity between the degree-corrected SBM of Karrer and Newman (2011) and our TDD-SBM is that at the MLE of the TDD-SBM, the expectation of the (time-aggregated) degree of a node i is equal to the degree of node i from the observed data. That is, $\sum_t \sum_j (\mu_{ijt} + \mu_{jit}) = k_i$. (See Appendix A.2 for the proof.) For our mixed-membership SBM, we are not aware of such a precise relationship between the data and the expected value of any model statistics.

We now substitute the above conditional parameter estimates into the unnormalized log-likelihood of the TDD-SBM to obtain

$$\begin{aligned} \sum_t \left[\sum_{i,j} \left(A_{ijt} \log \left(\frac{k_i}{\kappa_{g_i}} \right) + A_{jit} \log \left(\frac{k_j}{\kappa_{g_j}} \right) \right) + \sum_{g,h} m_{ght} \log m_{ght} - \sum_{g,h} m_{ght} \right] \\ = \sum_i k_i \log k_i - \sum_i k_i \log \kappa_{g_i} + \sum_t \sum_{g,h} m_{ght} \log m_{ght} - \tilde{m}, \end{aligned}$$

where \tilde{m} is the total number of edges in the network. We do a similar calculation as one in Karrer and Newman (2011) and write

$$\begin{aligned} \sum_i k_i \log \kappa_{g_i} &= \sum_t \sum_g \sum_{i \in g} k_{it} \log \kappa_g \\ &= \sum_t \sum_g \sum_{i \in g} (k_{in,it} \log \kappa_g + k_{out,it} \log \kappa_g) \\ &= \sum_t \sum_g \kappa_{out,gt} \log \kappa_g + \sum_t \sum_h \kappa_{in,ht} \log \kappa_h \\ &= \sum_t \sum_g \sum_h m_{ght} \log \kappa_g + \sum_t \sum_g \sum_h m_{ght} \log \kappa_h \\ &= \sum_t \sum_{g,h} m_{ght} \log \kappa_g \kappa_h, \end{aligned}$$

where $k_{in,it}$ and $k_{out,it}$ are the in-degrees and out-degrees of node i in layer t , the quantity $\kappa_{in,gt} = \sum_{i \in g} k_{in,it}$ is the number of edges that arrive at g in hour t , and $\kappa_{out,gt} = \sum_{i \in g} k_{out,it}$ is the number of edges that depart from g in hour t . Including only the terms that depend on block assignments yields the objective function

$$\sum_t \sum_{g,h} m_{ght} \log \left(\frac{m_{ght}}{\kappa_g \kappa_h} \right), \tag{7}$$

which we maximize to obtain the MLE of the block assignments.

Unlike the directed SBM of Zhu *et al.* (2013), our TDD-SBM does not have two degree-correction parameters for each station. Their directed SBM includes a parameter θ_i^{in} to modify in-degree and a parameter θ_i^{out} to modify out-degree. The validity of using undirected degree-correction parameters $\{\theta_i\}_{i \in \mathcal{N}}$ to model directed networks depends on the presence of a large correlation between the time-aggregated in-degrees and time-aggregated out-degrees of the nodes. In this situation, including distinct θ_i^{in} and θ_i^{out} parameters would lead to overfitting. The Pearson

correlations between in-degree and out-degree are larger than 0.9 for all of our bicycle-sharing networks (see Section 2). To increase the applicability of our work to other networks (which may not have such large correlations), we provide estimation code in our implementation materials (see Section 4.2) for a version of our TDD-SBM with directed degree-correction parameters θ_i^{in} and θ_i^{out} .

4. Computations

In this section, we describe the algorithms that we use to estimate the parameters of our TDMM-SBM and our TDD-SBM.

4.1 Inference using the TDMM-SBM

Let $\omega = \{\omega_{ght}\}$ be the $K \times K \times T$ array of block-connectivity parameters, and let $\mathbf{C} = \{C_{ig}\}$ be the matrix of node-strength parameters. We estimate the model parameters using a two-step gradient descent.⁴ First, we move in the direction of the gradient with respect to ω and update the block-connectivity parameters. Second, we move along the direction of the gradient with respect to \mathbf{C} and update the node-strength parameters.

In our algorithm, $\omega^{(n)}$ and $\mathbf{C}^{(n)}$ denote the n^{th} updates of the block-connectivity and node-strength parameters, respectively. We initialize the algorithm with random values $\omega^{(0)}$ and $\mathbf{C}^{(0)}$ with entries that are distributed according to $\exp(X)$, where X is a Gaussian random variable with mean 0 and variance 1. (That is, we draw random values from a log-normal distribution.) We denote the mean activity along edge (i, j) with initial parameters $\omega^{(0)}$ and $\mathbf{C}^{(0)}$ by $\mu_{ijt}^{(0)}$. We scale the initial parameters so that the TDMM-SBM at the starting point of the optimization procedure has the same mean total number of trips as the data. Specifically, we multiply the block-connectivity parameters $\omega_{ght}^{(0)}$ by $\left(\sum_t \omega_{ght}^{(0)}\right)^{-1} \left(\sum_{i,j,t} A_{ijt}\right) / K^2$ and normalize \mathbf{C}^0 to satisfy the constraint $\sum_i C_{ig}^{(0)} = 1$ for each block g . This yields $\sum_{ijt} \mu_{ijt}^{(0)} = \sum_{i,j,t} \sum_{g,h} C_{ig}^{(0)} \omega_{ght}^{(0)} C_{jh}^{(0)} = \sum_{g,h,t} \omega_{ght}^{(0)} = \sum_{g,h} \left(\sum_{i,j,t} A_{ijt}\right) / K^2 = \sum_{i,j,t} A_{ijt}$, which is the total number of edges in the network. Without this scaling, the initial parameters can have very small magnitudes, such that the mean total number of trips from the TDMM-SBM with these initial parameters is much smaller than the total number of trips in the data. In that situation, the initial likelihood is much smaller than the likelihood at the MLE.

To ensure that our estimated parameters are nonnegative, we use the following change of variables: $\exp(\tilde{\omega}^{(n)}) = \omega^{(n)}$ and $\exp(\tilde{\mathbf{C}}^{(n)}) = \mathbf{C}^{(n)}$, where the exponential of a matrix or a vector signifies componentwise exponentiation of each of its entries. We can then write the gradient descent as

$$\begin{aligned} \tilde{\omega}^{(n+1)} &= \tilde{\omega}^{(n)} + \eta^{(n)} \nabla_{\omega} \ell \left(\mathbf{C}^{(n)}, \omega^{(n)} \right) \exp \left(\tilde{\omega}^{(n)} \right), \\ \tilde{\mathbf{C}}^{(n+1)} &= \tilde{\mathbf{C}}^{(n)} + h^{(n)} \nabla_{\mathbf{C}} \ell \left(\mathbf{C}^{(n)}, \omega^{(n+1)} \right) \exp \left(\tilde{\mathbf{C}}^{(n)} \right), \end{aligned}$$

where $h^{(n)}$ and $\eta^{(n)}$ are small positive numbers. From the definitions of $\tilde{\mathbf{C}}^{(n)}$ and $\tilde{\omega}^{(n)}$, we write

$$\begin{aligned} \omega^{(n+1)} &= \omega^{(n)} \exp \left(\eta^{(n)} \nabla_{\omega} \ell \left(\mathbf{C}^{(n)}, \omega^{(n)} \right) \omega^{(n)} \right), \\ \mathbf{C}^{(n+1)} &= \mathbf{C}^{(n)} \exp \left(h^{(n)} \nabla_{\mathbf{C}} \ell \left(\mathbf{C}^{(n)}, \omega^{(n+1)} \right) \mathbf{C}^{(n)} \right). \end{aligned}$$

Let $h^{(0)} = \eta^{(0)} = \Delta > 0$ be the fixed initial step size. For our application, we choose $\Delta = 10^{-4}$. We generate two candidate updates for $\omega^{(n+1)}$ for the first step in our algorithm using $h^{(n+1)} = 1.2 h^{(n)}$

and $h^{(n+1)} = 0.8 h^{(n)}$, and we choose the one that gives the $\omega^{(n+1)}$ value that yields the larger $\ell(\mathbf{C}^{(n)}, \omega^{(n+1)})$. Similarly, we choose the one of $\eta^{(n+1)} = 1.2 \eta^{(n)}$ and $\eta^{(n+1)} = 0.8 \eta^{(n)}$ that gives the $\mathbf{C}^{(n+1)}$ value that yields the larger $\ell(\mathbf{C}^{(n+1)}, \omega^{(n+1)})$.

We compute the gradient of the log-likelihood function ℓ using the chain rule. Recall that we compute the log-likelihood in two parts. One part is the computation of the mean $\mu_{ijt} = \sum_{g,h} C_{ig} \omega_{gh}^t C_{jh}$ of the number of trips from node i to node j at time t . We then insert the expression for the mean into $\ell = \sum_t \sum_{i,j} (A_{ijt} \log(\mu_{ijt}) - \mu_{ijt})$. We compute the derivative of ℓ with respect to μ_{ijt} to obtain

$$\frac{\partial \ell}{\partial \mu_{ijt}} = \frac{A_{ijt}}{\mu_{ijt}} - 1.$$

The derivative of μ_{ijt} with respect to C_{kg} is

$$\frac{\partial \mu_{ijt}}{\partial C_{kg}} = \delta_{ki} \sum_h \omega_{gh} C_{jh} + \delta_{kj} \sum_h C_{ih} \omega_{hgt},$$

where δ_{ab} denotes the Kronecker delta (i.e., $\delta_{ab} = 1$ if $a = b$ and $\delta_{ab} = 0$ if $a \neq b$). The derivative of μ_{ijt} with respect to ω_{ght} is

$$\frac{\partial \mu_{ijt}}{\partial \omega_{ght}} = C_{ig} C_{jh}.$$

Using the above calculations, we see that the derivatives of ℓ with respect to C_{kg} and ω_{ght} are

$$\begin{aligned} \frac{\partial \ell}{\partial C_{kg}} &= \sum_{t=0}^{23} \sum_{i,j \in \mathcal{N}} \frac{\partial \ell}{\partial \mu_{ijt}} \frac{\partial \mu_{ijt}}{\partial C_{kg}} \\ &= \sum_{t=0}^{23} \left(\sum_{j \in \mathcal{N}} \left(\frac{A_{kjt}}{\mu_{kjt}} - 1 \right) \sum_h \omega_{gh} C_{jh} + \sum_{i \in \mathcal{N}} \left(\frac{A_{ikt}}{\mu_{ikt}} - 1 \right) \sum_h C_{ih} \omega_{hgt} \right), \\ \frac{\partial \ell}{\partial \omega_{ght}} &= \sum_{i,j \in \mathcal{N}} \frac{\partial \ell}{\partial \mu_{ijt}} \frac{\partial \mu_{ijt}}{\partial \omega_{ght}} \\ &= \sum_{i,j \in \mathcal{N}} \left(\frac{A_{ijt}}{\mu_{ijt}} - 1 \right) C_{ig} C_{jh}. \end{aligned}$$

We run gradient descent until four significant digits of the base-10 floating-point representation of the log-likelihood (3) do not change for 600 consecutive steps. For the networks that we examine, this usually takes between 600 and 5,000 iterations, with models with more blocks generally needing more iterations to reach this stopping criterion. Because of the nonconvexity of the log-likelihood function (3), we are not guaranteed to reach a global optimum. Most of the time, our method converges to an interesting local optimum (which may also be a global optimum) that appears to reveal functional roles of bicycle-sharing stations (see Section 6). Our results produce recognizable block-connectivity parameters ω_{ght} (e.g., reflecting home-work commute patterns and leisure patterns) and the parameters C_{ig} indicate known spatial divisions of the stations (e.g., residential versus commercial districts). In some cases, however, our algorithm converges to an uninteresting local optimum; for example, sometimes the block-assignment parameters C_{ig} for each station appear as if they were assigned independently at random. To improve our results, we run our algorithm several times (specifically, 10 times for each network) and store the estimate with the largest likelihood.

For several examples, we compare the parameters that we obtain from gradient descent to those that we obtain by running a Hamiltonian Monte Carlo (HMC) sampling method in a Bayesian framework with weak priors. (We implement this sampling method in Stan (Carpenter *et al.*, 2017; Stan Development Team, 2018).) The log-likelihoods that result from our gradient-descent method are as good or better than those that we obtain with the HMC method. The HMC method is more computationally intensive and memory intensive than our gradient-descent method, although it may be preferable in applications in which one has meaningful prior information about parameters. Improving our optimization method and investigating trade-offs between accuracy and efficiency are worthwhile topics for future work. For instance, it likely will be beneficial to adapt optimization methods for related time-dependent SBMs (Xing *et al.*, 2010; Ho *et al.*, 2011; Yang *et al.*, 2011; Xu and Hero, 2014; Matias *et al.*, 2018) to the optimization of our SBMs.

The repository at https://github.com/jcarlen/tdsbm_supplementary_material has our Python implementation of our gradient-descent method and code for our inference using Stan in R.

4.2 Inference using the TDD-SBM

To fit our TDD-SBM, we use a Kernighan–Lin-type (KL-type) algorithm (Kernighan and Lin, 1970) that we base on the one in Karrer and Newman (2011). (Karrer and Newman (2011) noted for their time-independent degree-corrected SBM that their KL-type algorithm is faster than node-switching algorithms and that it achieves results that are similar to those from such algorithms.) Given K blocks, we initialize the KL-type algorithm by assigning each node to a block uniformly at random, so each node has probability $1/K$ of being assigned to a given block. The algorithm then calculates the best possible block reassignment for any single node with respect to the associated change in log-likelihood (either the largest increase or the smallest decrease). It then chooses a new node uniformly at random and determines its best reassignment with respect to change in log-likelihood. The algorithm cycles through all nodes; a key feature of the algorithm is that a node that has been reassigned cannot change blocks again until all other nodes have been reassigned. A sequential reassignment of all nodes constitutes one step of the algorithm. The algorithm then searches all of the states that occurred during the step, and it selects the state with the maximum log-likelihood of any state in that step. This state is the starting point of the next step of the algorithm. A single run of the algorithm finishes when a step does not increase the log-likelihood beyond a preset tolerance value near 0. (In practice, we use 1×10^{-4} .) To find block assignments that are as good as possible, we do many runs of the algorithm for each network; this process is easily parallelizable. In our examples, we use 50 runs per network. We initialize each run randomly as described above.

Another key feature of the KL-type algorithm is that changes in the block memberships of nodes affect only the terms of the objective function that involve the origin and destination blocks of the changes. The simplified objective function (7) is a sum over block-pair terms over T time layers. Consequently, we do not need to recalculate the full objective function for each node reassignment. This estimation method is practical for networks with up to a few thousand nodes and a small number of blocks, depending on available computing resources.

We implement our KL-type algorithm for our TDD-SBM in an R package using **Rcpp** (Eddelbuettel and François, 2011; R Core Team, 2018). The back-end calculations are in C++ for speed, and we return results in R to enable visualization and other downstream tasks. One can also use our R package to estimate the parameters of time-independent SBMs (including directed and/or degree-corrected ones), a variant of our TDD-SBM without degree-correction, and a variant of our TDD-SBM with directed degree-correction parameters θ_i^{in} and θ_i^{out} . This facilitates comparisons of inferences from various related SBMs. See <https://github.com/jcarlen/sbm> for our R package **sbmt** for parameter estimation for our TDD-SBM.

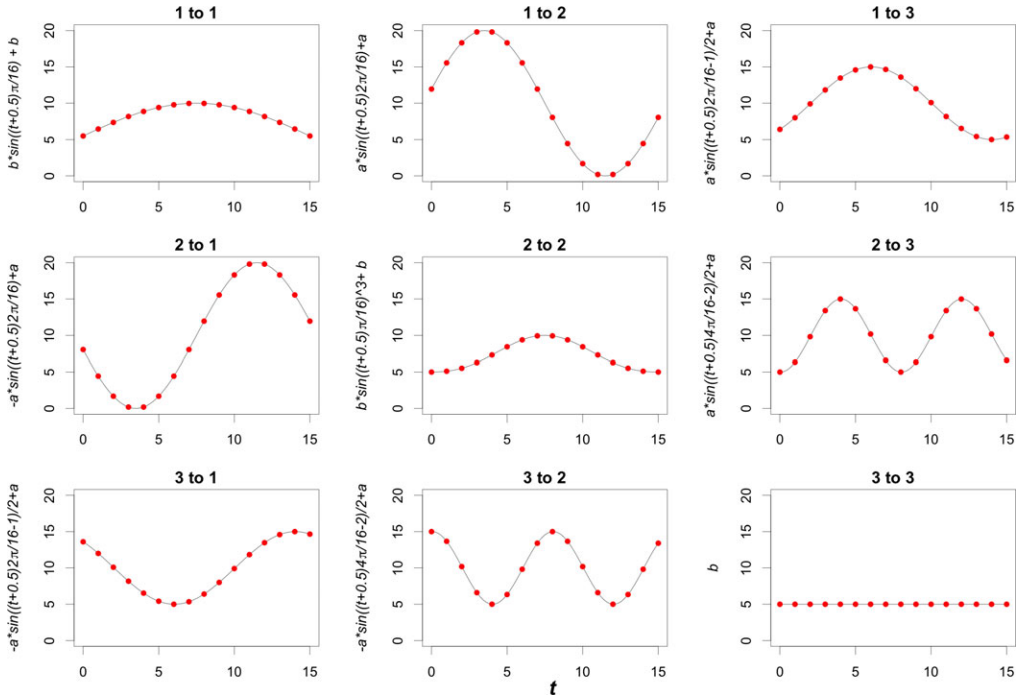


Figure 2. In the TDD-SBM networks that we construct, the red dots indicate the mean weights of the edges from block g to block h in time layer t , where $g, h \in \{1, 2, 3\}$ and $t \in \{0, 1, \dots, 15\}$. The title of each panel indicates the blocks g and h (e.g., “1 to 2”).

5. Evaluation of model fitting for synthetic networks

In this section, we generate networks using three models: our TDD-SBM, a variant of our TDD-SBM without degree-correction, and our TDMM-SBM. Unless we specify otherwise, we use the term TDD-SBM to refer to our time-dependent discrete-membership SBM with degree-correction. For networks that we generate using discrete-membership models, we estimate model parameters using the TDD-SBM, the TDD-SBM variant without degree-correction, and the Poisson-process SBM (PPSBM) of Matias *et al.* (2018),⁵ which also does not have degree-correction. For the networks that we generate using our TDMM-SBM, we estimate model parameters using our TDMM-SBM and our TDD-SBM. We compare how well these estimated models fit the generated data. We illustrate three key points. First, our estimation methods are able to recover data-generating parameters for generated examples. Second, incorporating degree-correction in our models facilitates the identification of blocks with similar temporal activity patterns, as opposed to ones with similar activity levels (i.e., similar degrees), when nodes have heterogeneous degrees. Third, fitting our discrete-membership model (TDD-SBM) to networks that are generated by our mixed-membership model (TDMM-SBM) can result in smaller likelihoods of the observed data than when fitting to these data using our TDMM-SBM. That is, one can lose information when modeling mixed-membership networks as discrete-membership networks.

We now discuss how we generate TDD-SBM networks. We construct networks with 16 time layers to facilitate comparison to the PPSBM. In Figure 2, we show the mean edge weights between nodes in the block pairs for each time layer of the TDD-SBM. We design the underlying curves to resemble patterns that we observed in the urban bicycle-sharing networks. Using these mean edge weights and a specified number of nodes in each block, we calculate the expected sums of the edge weights from one block to another for each layer. In other words, we calculate the MLE $\hat{\omega}$ (see (5)), given this information. We set the parameter $\omega = \{\omega_{ght}\} = \hat{\omega}$ to generate our networks.

We construct both two-block and three-block networks, where the former only uses values in the upper 2×2 panel of Figure 2. For simplicity, we divide block membership equally between blocks in the TDD-SBM. For example, if there are 30 nodes and two blocks, 15 nodes belong to each block. We assume that the degree-correction parameters θ take one of five distinct values $\{1x, 2x, \dots, 5x\}$, which are distributed evenly within each block; the value of x satisfies the blockwise sum constraints on θ .

Using the above setup, we generate multilayer networks from our TDD-SBM. We also generate and fit networks using a version of our TDD-SBM without degree-correction. A discussion of this model is beyond the scope of this paper, but we implemented parameter estimation for it in our `sfmt` package. The log-likelihood of this model is

$$\ell(G; \omega) = \sum_t \sum_{g,h} \sum_{i \in g, j \in h} [A_{ijt} \log(\omega_{ght}) - \omega_{ght} - \log(A_{ijt}!)] ,$$

which is the log-likelihood of the TDD-SBM without the degree-correction parameter θ . For each experiment, we produce 100 instantiations of the networks. All of the networks that we generate are directed, and we allow self-edges.

For the networks that we generate using the TDD-SBM or the TDD-SBM without degree-correction, we estimate model parameters by fitting them using our TDD-SBM, our TDD-SBM without degree-correction, and the PPSBM of Matias *et al.* (2018) (which is implemented in the `ppsbm` package in R (Giorgi *et al.*, 2018)). Their PPSBM is a time-dependent SBM that is similar to our TDD-SBM without degree-correction. It differs in that it assumes continuous-time networks and either adaptively splits data into equal-sized bins (to use a nonparametric histogram-based estimator) or uses a kernel estimator of time-dependent intensity functions. For our comparison, we implement the former approach because the data input structure aligns well with our preprocessed data. We estimate the parameters of the PPSBM using the variational expectation-maximization (EM) algorithm from the `ppsbm` package. The package includes a method to select the number of blocks using integrated classification likelihood (ICL).

In Table 1, we show the results for networks that we generate using our TDD-SBM and our TDD-SBM without degree-correction. The first two columns give the number of blocks (K) and number of nodes (N) that we use to construct the networks. (Recall that we predetermine these values in our models, so they are given.) We indicate in the “Gen DC” column whether or not the data-generating TDD-SBM has degree-correction. We indicate the model type that we use to fit the data in the “Fit Method” column, and we indicate whether or not it incorporates degree-correction in the “Fit DC” column.

We evaluate the success of each estimation method at detecting the true block assignments by calculating the adjusted Rand index (ARI) using the `fossil` package in R (Vavrek, 2011). We report the mean ARI values of 100 network instantiations in the “ARI” column of Table 1. The ARI is a function of two group-membership vectors and achieves its maximum value of 1 when they match perfectly (up to label permutations). We expect an ARI of 0 when we compare two uniformly random permutations of block assignments that preserve the block sizes.

We evaluate the ability of the estimation methods to recover the true value of the parameter ω using the mean absolute-percentage error (MAPE) between $\hat{\omega}$ and ω . The formula for the MAPE is $\text{MAPE}(\hat{\omega}) = \frac{1}{ght} \sum_{ght} \left| \frac{\hat{\omega}_{ght} - \omega_{ght}}{\omega_{ght}} \right|$; we report the mean MAPE values of 100 network instantiations in the “MAPE($\hat{\omega}$)” column of Table 1. We also compare the unnormalized log-likelihoods of the generated networks under the data-generating model (we show their means in the “Gen LLIK” column) to the unnormalized log-likelihoods under the estimated model. Given the similarities of the candidate models, this yields an overall measure of how well we estimate the means of the Poisson distributions that describe the edge weights. We show the means of the differences between these two quantities in the “Diff LLIK” column, where positive values signify that the data are more likely under the estimated model.

Table 1. Evaluation of parameter-estimation methods for networks that we generate using our TDD-SBM and our TDD-SBM without degree-correction. We estimate parameters using both of these models and the PPSBM of Matias *et al.* (2018). For each experiment, we indicate the number of blocks (K) and the number of nodes (N) in the networks. The “Gen DC” column indicates whether or not we use degree-correction when generating the networks, and the “Fit DC” column indicates whether or not we use degree-correction when we fit the networks. The “Fit Method” column indicates which model we use for fitting. To the right of the vertical line, we show the means of several statistics over 100 instantiations of the networks, with the associated standard deviations in parentheses. “ARI” is the adjusted Rand index between the inferred block assignments and the ground-truth assignments. The quantity $\text{MAPE}(\hat{\omega})$ is the mean absolute-percentage error, expressed as a fraction, between the block-connectivity parameters $\hat{\omega}$ and ω . “Gen LLIK” is the unnormalized log-likelihood of a generated network under the data-generating model. “Diff LLIK” is the result of subtracting the unnormalized log-likelihood of a generated network under the data-generating model from its unnormalized log-likelihood under the estimated model

K	N	Gen DC	Fit DC	Fit Method	ARI	$\text{MAPE}(\hat{\omega})$	Gen LLIK	Diff LLIK
2	30	F	F	TDD-SBM	1 (0)	0.03 (0)	174,414 (713)	32 (6)
3	30	F	F	TDD-SBM	1 (0)	0.03 (0)	169,917 (784)	70 (9)
2	30	T	F	TDD-SBM	1 (0)	0.03 (0)	204,808 (952)	−30,390 (221)
3	30	T	F	TDD-SBM	0.67 (0)	1.80 (0.02)	200,581 (878)	−16,673 (151)
2	30	F	T	TDD-SBM	1 (0)	0.03 (0)	174,467 (861)	45 (7)
3	30	F	T	TDD-SBM	1 (0)	0.03 (0)	169,776 (855)	86 (9)
2	30	T	T	TDD-SBM	1 (0)	0.03 (0)	204,964 (941)	32 (6)
3	30	T	T	TDD-SBM	1 (0)	0.03 (0)	200,687 (1,022)	71 (8)
2	30	F	F	PPSBM	1 (0)	0.03 (0)	174,475 (827)	31 (5)
3	30	F	F	PPSBM	1 (0)	0.03 (0)	169,928 (983)	59 (11)
2	30	T	F	PPSBM	0.04 (0.06)	2.71 (0.36)	204,907 (887)	−31,342 (310)
3	30	T	F	PPSBM	0.16 (0.04)	1.05 (0.22)	200,551 (899)	−18,325 (951)
2	90	F	F	TDD-SBM	1 (0)	0.01 (0)	1,569,720 (2,702)	33 (6)
3	90	F	F	TDD-SBM	1 (0)	0.01 (0)	1,529,181 (2,512)	72 (9)
2	90	T	F	TDD-SBM	1 (0.04)	0.04 (0.29)	1,843,773 (2,548)	−273,917 (993)
3	90	T	F	TDD-SBM	0.70 (0)	1.80 (0.01)	1,805,703 (2,600)	−150,714 (479)
2	90	F	T	TDD-SBM	1 (0)	0.01 (0)	1,569,907 (2,851)	76 (8)
3	90	F	T	TDD-SBM	1 (0)	0.01 (0)	1,528,897 (2,574)	115 (10)
2	90	T	T	TDD-SBM	1 (0)	0.01 (0)	1,843,215 (2,921)	32 (5)
3	90	T	T	TDD-SBM	1 (0)	0.01 (0)	1,806,107 (3,201)	74 (9)
2	90	F	F	PPSBM	1 (0)	0.01 (0)	1,569,564 (2,489)	32 (5)
3	90	F	F	PPSBM	1 (0)	0.01 (0)	1,528,835 (2,333)	63 (9)
2	90	T	F	PPSBM	0.01 (0)	2.77 (0.01)	1,843,826 (3,099)	−281,730 (755)
3	90	T	F	PPSBM	0.20 (0)	1.10 (0.1)	1,806,072 (2,952)	−167,642 (3,864)

We find perfect matches (i.e., an ARI of 1 with standard deviation of 0) between the true and estimated block-membership assignments whenever the network-generation and estimation methods match or the estimation model (TDD-SBM) is a generalization of the network-generation model (TDD-SBM without degree-correction). We also find perfect matches when the networks are generated by our TDD-SBM without degree-correction and are estimated by the PPSBM; this illustrates the similarity between these two models. In all of those cases, there is very little error in $\hat{\omega}$, as indicated by the small MAPE values, and the likelihoods of the generated networks under the estimated model are slightly larger than the likelihoods under the data-generating

Table 2. Evaluation of parameter-estimation methods for networks that we generate using our TDMM-SBM. For each experiment, we indicate the number of blocks (K) and the number of nodes (N) in the networks. To the right of the vertical line, we show the means of several statistics over 100 instantiations of the networks, with associated standard deviations in parentheses. “BAE” is the blockwise absolute error of the block-membership parameters. The quantity $\text{MAPE}(\hat{\omega})$ is the mean absolute-percentage error, expressed as a fraction, between the block-connectivity parameters $\hat{\omega}$ and ω . We show “-” in this column when the result is extremely large due to values of ω_{ght} near 0. The quantity $\text{MAPE}_p(\hat{\omega})$ is the pairwise MAPE of $\hat{\omega}$. “Gen LLIK” is the unnormalized log-likelihood of a generated network under the data-generating TDMM-SBM. “Diff LLIK” is the result of subtracting the unnormalized log-likelihood of a generated network under the data-generating model from its unnormalized log-likelihood under the estimated TDMM-SBM. “Diff LLIK Discrete” is the result of subtracting the unnormalized log-likelihood of a generated network under the data-generating model from its unnormalized log-likelihood under the degree-corrected TDD-SBM

K	N	BAE(\hat{C})	MAPE($\hat{\omega}$)	MAPE _p ($\hat{\omega}$)	Gen LLIK	Diff LLIK	Diff LLIK Discrete
2	30	0.02 (0.004)	0.10 (0.01)	0.04 (0.01)	198,447 (969)	58 (8)	-2,243 (58)
3	30	0.22 (0.054)	0.74 (0.21)	0.52 (0.10)	168,880 (729)	94 (23)	-241 (25)
2	90	0.02 (0.001)	0.09 (0.01)	0.02 (0.01)	1,651,845 (2,806)	117 (11)	-22,830 (195)
3	90	0.21 (0.077)	0.48 (0.19)	0.39 (0.09)	1,581,507 (2,343)	135 (109)	-4,221 (82)
2	61	0.05 (0.004)	-	0.08 (0.01)	-37,351 (233)	105 (10)	-597 (38)
3	61	0.06 (0.005)	-	0.29 (0.03)	-35,214 (211)	168 (11)	-1,569 (54)

model. (Recall that we fit the former to the generated networks.) This indicates that our TDD-SBM estimation method is effective in these examples.

When we generate networks using our TDD-SBM with degree-correction, the two methods without degree-correction (i.e., the TDD-SBM variant and PPSBM) do a poor job of fitting the data. In these cases, we obtain large values of both ARI and MAPE for three-block networks. Even in the two-block examples, for which we recover the blocks effectively, we obtain much smaller log-likelihoods of the generated networks under their estimated models than under the data-generating model. This is the case because using a model without degree-correction implies that all edges between a given pair of blocks are identically distributed, regardless of the nodes to which the edges are attached. This does not accurately reflect our generated networks, which have heterogeneous degrees within blocks.

For estimation using PPSBM, the results in Table 1 reflect the fact that we set the number of blocks to be equal to the value (which is two or three in our examples) in our data-generating model. When we apply the block-selection method of Matias *et al.* (2018) to models with between one and four blocks (using the implementation in the `ppsbm` package), we identify the true number of blocks for models without degree-correction. However, when we apply it to networks that we generate using our TDD-SBM with degree-correction and let it search for up to four blocks, we overestimate the number of blocks. The PPSBM method attempts to classify nodes based both on their roles and on their activity levels, so it requires extra blocks.

We generate TDMM-SBM networks using the same values of ω , with corresponding numbers of blocks and nodes, that we used to generate TDD-SBM networks. We are interested in how well our mixed-membership model (i.e., TDMM-SBM) is able to recover data-generating parameters and how much log-likelihood we lose when we use a TDD-SBM to fit a network that we generate using the TDMM-SBM. For each combination of number of blocks (K) and number of nodes (N), we draw values of C uniformly at random from $\{0, 1, \dots, 5\}$ and then normalize each block so that $\sum_i C_{ig} = 1$. We also include results for networks that we generate with our TDMM-SBM using parameters that we determine by fitting to the Los Angeles bicycle-sharing network (see Section 6.1) with two and three blocks. In these examples, $N = 61$ because that is the number bicycle-sharing stations in that network. As with the TDD-SBM, we produce 100 network instantiations for each case. We summarize our results in Table 2.

Because block membership is mixed, we do not use ARI to compare \hat{C} with C . Instead, we calculate the blockwise absolute error (BAE), which is the sum of the absolute errors divided by the number of blocks. That is, $\text{BAE}(\hat{C}) = \frac{1}{K} \sum_{i,g} |\hat{C}_{ig} - C_{ig}|$. To evaluate the error in $\hat{\omega}$, we calculate

the MAPE. We also calculate a “pairwise” MAPE of $\hat{\omega}$ (which we denote by $\text{MAPE}_p(\hat{\omega})$ in Table 2), in which we average the absolute percentage error over each pair of blocks. That is, $\text{MAPE}_p(\hat{\omega}) = \frac{1}{K^2} \sum_{g,h} \frac{\sum_t |\hat{\omega}_{ght} - \omega_{ght}|}{\sum_t \omega_{ght}}$. The pairwise MAPE is less susceptible than MAPE to having very large values when individual values of ω_{ght} are near 0.

We compare the log-likelihoods of the generated networks under the data-generating TDMM-SBM (“Gen LLIK”) to their log-likelihoods under the estimated parameters, and we report the means of the differences (the latter minus the former) in the “Diff LLIK” column of Table 2. Finally, we fit our TDD-SBM to the generated networks and calculate the log-likelihoods of the networks under that model. We report the means of these log-likelihoods minus those that we averaged for the “Gen LLIK” column in the “Diff LLIK Discrete” column.

We find that the likelihoods of the generated networks when we use the estimated parameters are slightly larger on average than the likelihoods of the networks when we use the data-generating parameters. (Because of the randomness in the networks that we generate, we expect that a network that we generate using a given model will have a larger likelihood when we use the parameters that we fit to that network than when we use the data-generating parameters.) Additionally, the mean likelihoods of the generated networks under parameters that we estimate using our TDD-SBM are significantly smaller than those that we obtain under the data-generating TDMM-SBM parameters. (See the “Diff LLIK Discrete” column of Table 2.) In short, we gain a lot of information by employing our mixed-membership model in this case.

The BAE, MAPE, and MAPE_p values (i.e., the “error values”) tend to be small for the two-block models, indicating that we are recovering the data-generating parameters fairly well in these cases. (We do not report MAPE values for generated networks when we use the estimated parameters of our TDMM-SBM for Los Angeles, which has $N = 61$ bicycle-sharing stations, because a few values of ω that are very close to 0 drastically inflate those values.) However, the values of BAE, MAPE, and MAPE_p are much larger for the three-block models (except for the BAE for Los Angeles). The fact that the estimated parameters achieve similar likelihoods as the data-generating parameters despite fairly large values of BAE, MAPE, and MAPE_p suggests that our TDMM-SBM is not completely identifiable. It is important to be aware of this identifiability limitation when interpreting the parameters of our TDMM-SBM. This is especially true for ω , given the large MAPE and MAPE_p values for all three-block TDMM-SBMs.

The identifiability of time-dependent SBMs is a relatively unexplored research topic. Matias and Miele (2017) pointed out that the time-dependent SBM of Xu and Hero (2014) is not fully identifiable, and they introduced a demonstrably identifiable (up to label permutations) time-dependent SBM. They also established that identifiability is an inherent limitation of time-dependent SBMs in which block-membership parameters and block-connectivity parameters vary with time, unless one applies additional constraints or penalization. As we described in Section 3, our models allow only the block-connectivity parameters to vary. Further research is necessary to establish what conditions ensure the identifiability of our time-dependent SBMs.

6. Results on bicycle-sharing networks

We apply our models to bicycle-sharing networks in Los Angeles, San Francisco, and New York City. (See Section 2 for full descriptions of these data sets.) The networks that we examine in downtown Los Angeles and San Francisco are relatively small; they have 61 stations and 35 stations, respectively. The stations in these networks are concentrated in downtown areas, where residential buildings and high-rise office buildings are interspersed. The New York City network is much larger than the other two networks. It has about 600 stations, which encompass a range of commercial areas, residential neighborhoods, parks, and manufacturing areas.

Before discussing our results in detail, we highlight some of our key insights. We present results for two-block models for the Los Angeles (see Section 6.1) and San Francisco (see Section 6.2)

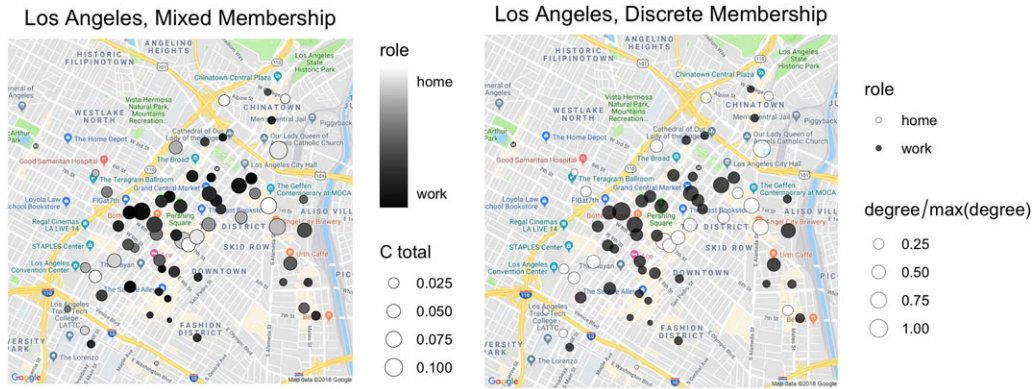


Figure 3. Classification of bicycle-sharing stations in downtown Los Angeles using (left) a two-block TDMM-SBM and (right) a two-block TDD-SBM. The sizes of the nodes take continuous values. In the left panel, we scale the size of each node based on its value of $\sum_g C_{ig}$. In the right panel, we scale the size of each node based on its degree.

bicycle-sharing networks that illustrate the ability of our models to detect functional roles in the form of “home” and “work” roles of bicycle-sharing stations. These results enable us to compare bicycle-sharing systems across cities. We fit discrete-membership and mixed-membership models with two, three, and four blocks to each of our data sets; we also fit five-block and six-block models in several cases. We provide the outputs of these models and all code to recreate our results at https://github.com/jcarlen/tdsbm_supplementary_material. In most cases, we find that additional blocks do not illuminate new functional roles beyond the home and work roles. For example, using models with additional blocks for Los Angeles and San Francisco results in small variations of these roles but increases the noise in the block-connectivity parameters $\hat{\omega}_{ght}$. In Section 6.3, we present five-block TDMM-SBM and TDD-SBM fits of a subset of the New York City bicycle-sharing network that illustrate another role (a “park” block) that does not appear distinctly when we employ models with fewer blocks. A method to select the number of blocks is beyond the scope of our paper, but we discuss aspects of this model-selection problem in Appendix A.4 and illustrate it by fitting mixed-membership models with up to 10 blocks on the Los Angeles bicycle-sharing network.

6.1 Downtown Los Angeles

In Figure 3, we show the mixed-membership (TDMM-SBM) and discrete-membership (TDD-SBM) block assignments of two-block models for the downtown Los Angeles bicycle-sharing network. For the TDMM-SBM, we scale the size of each node i in our plots based on its value of $\sum_g C_{ig}$. We refer to these sums as “C total” values. These values correlate strongly with node degree (specifically, with the sum of the in-degrees and out-degrees of all layers), which is evident in the similarity of node sizes in the left and right panels of Figure 3. For both models, we observe that home and work blocks are interspersed geographically. As part of our discussion in this subsection, we will describe our method for determining the block labels in Figure 3.

The TDMM-SBM output reveals a group of stations in the left panel of Figure 3 that are neither strongly home-identified nor strongly work-identified. Instead, they possess a roughly even mixture of the two types. For this network, the TDD-SBM output is very similar to what we obtain from a discretization of the TDMM-SBM output (where we discretize by assigning each node i to the block with the maximum value of its C_{ig} parameter), but this is not true for all of our bicycle-sharing networks.

Our model does not yield “home” and “work” labels for each block on its own, so we use the time-dependent block-connectivity parameter estimates $\hat{\omega}_{ght}$ to assign these labels. We assign the

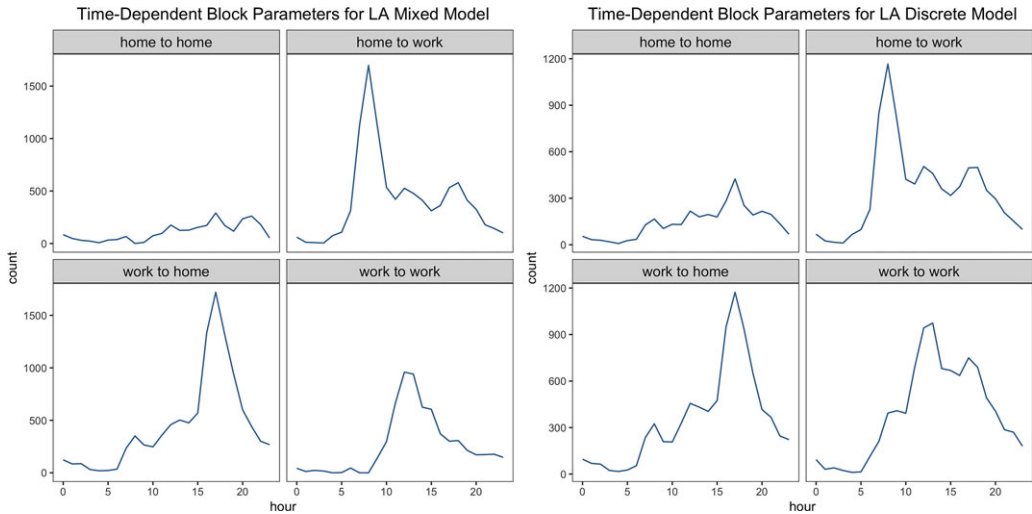


Figure 4. Estimated time-dependent block-connectivity parameters $\hat{\omega}_{ght}$ for (left) the two-block TDM-SBM and (right) the two-block TDD-SBM for the downtown Los Angeles bicycle-sharing network.

labels heuristically under the assumption that the “home” block is the origin of many trips to the work block in the morning and the “work” block is the origin of many trips to the home block in the evening. Figure 4, which shows $\hat{\omega}_{ght}$ for each possible value of g and h , with the hour t on the horizontal axis, supports our labeling. Given our labeling, we observe a clear peak in home-to-work traffic in the mornings and work-to-home traffic in the evenings. We make similar “home” and “work” assignments for San Francisco and New York City. In Los Angeles, the traffic within the work block peaks in the middle of the day. This may represent lunchtime errands, leisure activity, or touristic activity. (There are many tourist attractions in the downtown area.) The traffic from the home block to itself has a mild evening peak and is the lowest overall among the block pairs.

To further evaluate our block labels, we use the zoning map for downtown Los Angeles from the Los Angeles Department of City Planning (Department of City Planning, City of Los Angeles, 2015).⁶ Zoning ordinances determine the allowable uses of city land. These ordinances distinguish between land that is available for commercial uses, industrial uses, residential uses, park districts, and other uses. In the background of Figure 5, we show a simplified version of the underlying zoning map in which we have grouped similar designations. The industrial areas include a mixture of manufacturing and commercial uses. Public facilities include government buildings, public schools, parking under freeways, and police and fire stations (Department of City Planning, City of Los Angeles, 2022). In downtown Los Angeles, manufacturing and industrial areas are split cleanly from residential areas, whereas residential and commercial areas are intermixed across the bicycle-sharing system’s coverage area.

Figure 5 illustrates that most stations that are strongly home-identified are in or near zones for pure residential use or mixed residential and commercial use. We find that many stations that are not predominantly home-identified or work-identified align with mixed-use residential/commercial zones. The discrete-role plot (see the right panel of Figure 3) has a stripe of “home” stations that cut diagonally through the “work” stations. In Figure 5, we see that this aligns roughly with areas that are zoned for purely residential use. By contrast, industrial and public-facility zones tend to have stations that are primarily work-identified, although some of the most strongly work-identified stations are in mixed-used areas.

One station that seems to deviate from the overall pattern is the heavily trafficked station at Union Station. Although it is adjacent to a public facility zone with many government buildings,

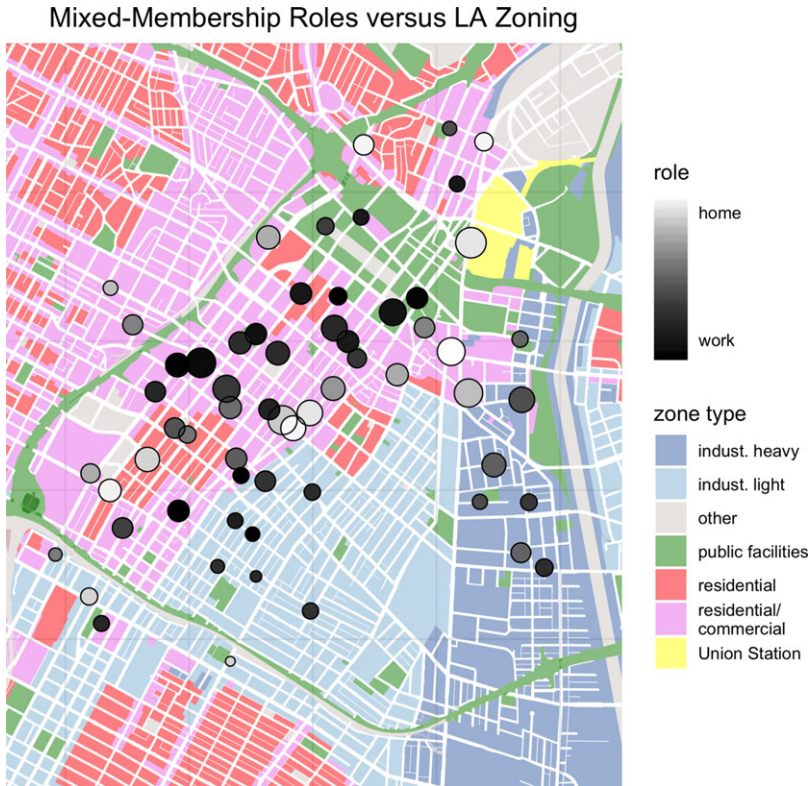


Figure 5. Two-block TDMM-SBM assignments of downtown Los Angeles bicycle-sharing stations overlaid on a simplified LA zoning map. The industrial zones include manufacturing and commercial areas. As in Figure 3, we scale the size of each node based on its value of $\sum_g C_{ig}$.

it is also strongly home-identified. This may seem surprising on its surface, but this classification is consistent with those of other home-identified stations because Union Station is a major transit hub for the Los Angeles metropolitan area. Many morning trips originate there as commuters transition from other modes of transportation, and many evening trips conclude there. This type of activity pattern is sensibly associated with home-identified stations. Such idiosyncrasies of transit hubs also arise in our results for San Francisco and New York City.

6.2 San Francisco

In Figure 6, we compare our two-block TDMM-SBM and two-block TDD-SBM for San Francisco. As we saw for Los Angeles, the San Francisco blocks are interspersed geographically and stations vary from strongly home-identified to strongly work-identified. The two most strongly home-identified stations are located at a major transit hub (the San Francisco Caltrain Station on 4th Street).

In Figure 7, we show the estimated traffic between the “home” and “work” blocks for the two-block TDMM-SBM and the two-block TDD-SBM. As in the downtown Los Angeles network, we observe inter-block commuting. However, unlike in downtown LA, the TDD-SBM reveals intra-block morning and evening peaks in both the home and work blocks. This may be due to last-mile commuting, such as using bicycle-sharing facilities to get to or from a transit station. The 10 most-popular trips within blocks (i.e., home-to-home travel or work-to-work travel when there are two blocks) that originate during the heavy commuting hours of 7–9 am and 4–6 pm all begin at major

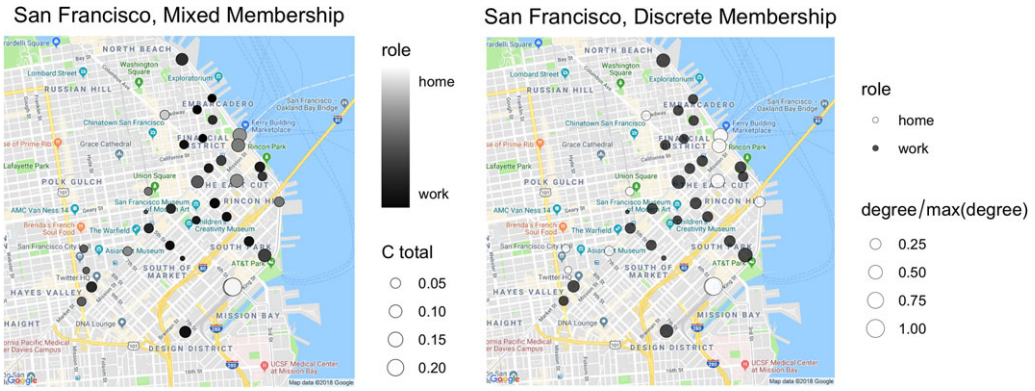


Figure 6. Classification of bicycle-sharing stations in San Francisco using (left) a two-block TDM-SBM and (right) a two-block TDD-SBM. The sizes of the nodes take continuous values. In the left panel, we scale the size of each node based on its value of $\sum_g C_{ig}$. In the right panel, we scale the size of each node based on its degree.

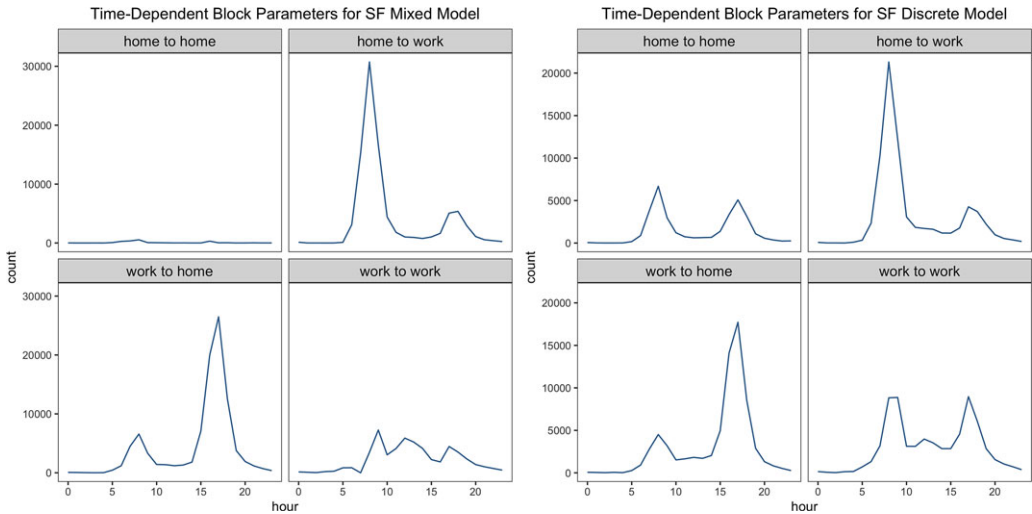


Figure 7. Estimated time-dependent block-connectivity parameters $\hat{\omega}_{ght}$ for (left) the two-block TDM-SBM and (right) the two-block TDD-SBM for the San Francisco bicycle-sharing network.

public transportation stops in the morning and end at those stops in the evening. Recognizing last-mile usage is important for integrating bicycle-sharing systems with nearby public transportation. One possible reason that we do not observe a similar phenomenon in downtown LA is that San Franciscans are more likely than residents of Los Angeles to use public transportation (Weikel and Karlamangla, 2015). The intra-block morning and evening peaks may also arise from the intermixing of commercial and residential uses of land. In other words, some travel within blocks may also constitute commuting.

6.2.1 Comparison to SBM variants

We briefly compare our above results for the Los Angeles and San Francisco bicycle-sharing networks to results that we obtain for these networks using a time-independent SBM and a time-dependent SBM without degree-correction. We do this to illustrate that both time-dependence

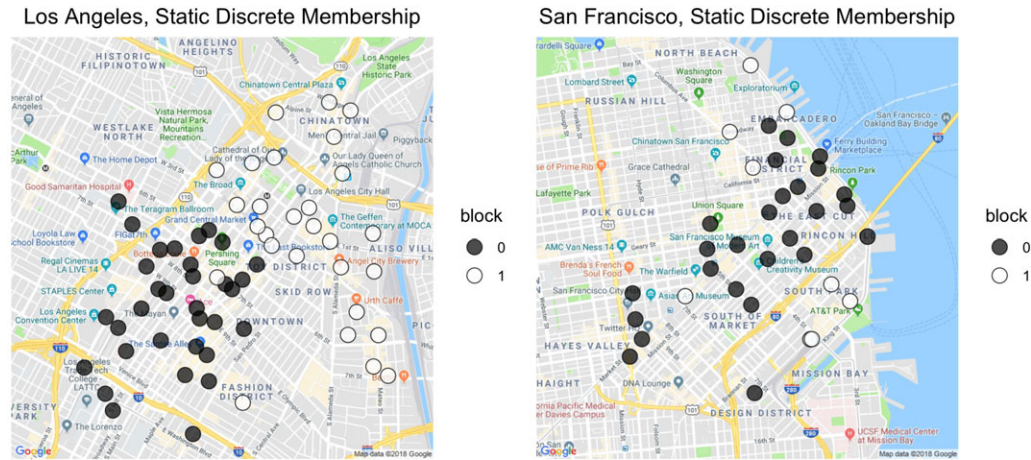


Figure 8. Estimated blocks from two-block TDD-SBMs for time-aggregated bicycle-sharing networks in (left) downtown Los Angeles and (right) San Francisco.

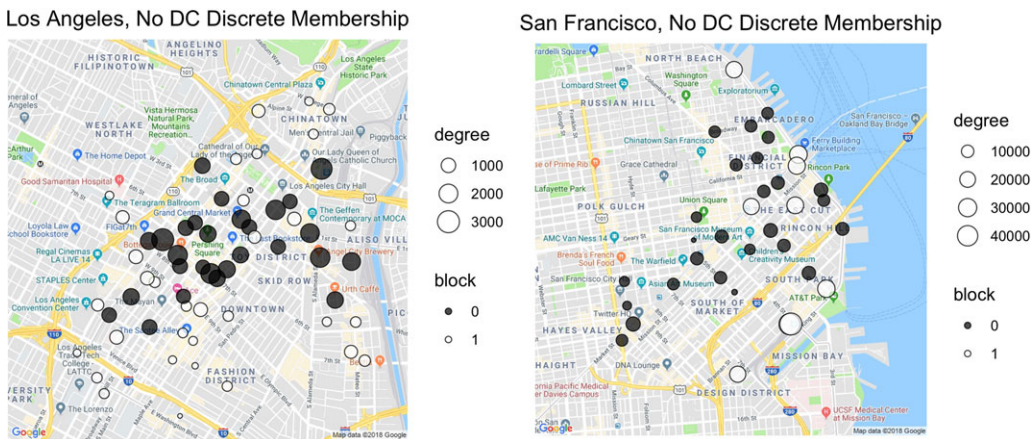


Figure 9. Estimated blocks from two-blocks TDD-SBMs without degree-correction for bicycle-sharing networks in (left) downtown Los Angeles and (right) San Francisco. We scale the size of each node based on its degree.

and degree-correction are important for achieving our goal of detecting functional roles of bicycle-sharing stations.

To fit the time-independent SBM, we construct the time-independent adjacency matrix with entries $\tilde{A}_{ij} = \sum_{t=0}^{23} A_{ijt}$ and fit a two-block TDD-SBM to the associated time-aggregated network with a single layer. For the downtown Los Angeles network, we observe a clear geographical division in the output of the time-independent SBM (see Figure 8). For the San Francisco network, the differences between the blocks of the time-dependent SBM and time-independent SBM are less noticeable, although they are still present. This confirms that our time-dependent SBMs detect usage patterns that are not evident in the time-aggregated data.

We now fit the variant of our two-block TDD-SBM without degree-correction (see Section 5) to the Los Angeles and San Francisco bicycle-sharing networks. In Figure 9, we see that stations are now divided into one group of small-degree stations and one group of large-degree stations, where the “degree” is the sum over all layers of the in-degrees and out-degrees. These blocks differ significantly from those that we obtained using the TDD-SBM with degree-correction (see Figures

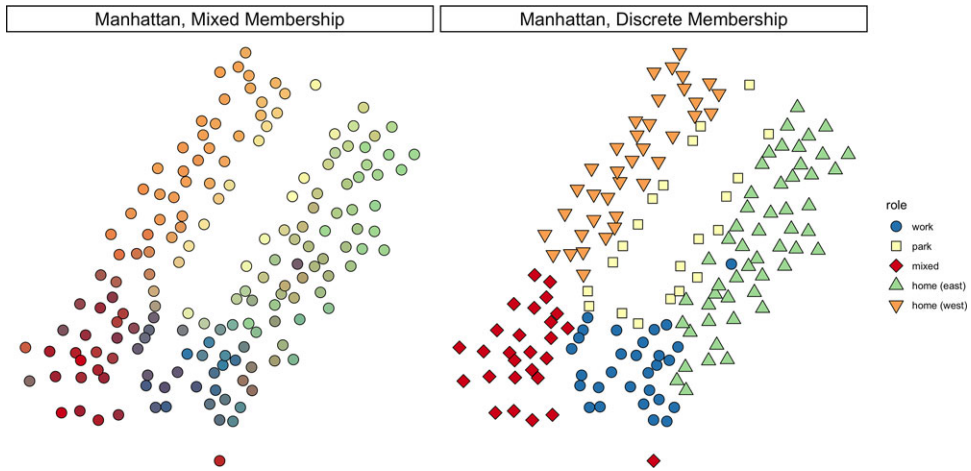


Figure 10. Comparison of the blocks that we infer in the Manhattan subnetwork (i.e., the “Manhattan (home)” block in the right panel of Figure A.4) of the New York City bicycle-sharing network using (left) a five-block TDMM-SBM and (right) a five-block TDD-SBM.

3 and 6). For example, for Los Angeles, the TDD-SBM with degree-correction classifies the high-traffic node at Union Station as a “home” station because of its frequent departures in the morning and frequent arrivals in the evening, whereas the TDD-SBM without degree-correction groups Union Station with other high-traffic stations that are classified predominantly as “work” stations in the degree-corrected TDD-SBM.

6.3 New York City

We fit our TDMM-SBM and our TDD-SBM to the New York City bicycle-sharing network with 2–6 blocks. Our results suggest that there is a limit to the size of networks for which our time-dependent SBMs can recover blocks that are based on functional roles of stations (rather than on geographical proximity). Therefore, in the present subsection, we apply our models to a subset of the New York City bicycle-sharing network to gain insight about the functional roles of stations in this subset. We select this subset based on the results of applying our three-block TDD-SBM to the entire New York City network. We examine the subnetwork that is induced by the stations that are in one of these three blocks. We explore three-block SBMs of the entire New York City network in Appendix A.3. In those results, we infer that stations are assigned to blocks along geographical lines, rather than to blocks with distinct functions.

6.3.1 Manhattan subnetwork

We examine the New York City bicycle-sharing network on a scale that is smaller than the full system and fit models to the subset of stations and trips that lie within the Manhattan (home) block of the three-block TDD-SBM (see Figure A.4) that we examine in Appendix A.3. We refer to this subnetwork as the “Manhattan” network, even though it does not include all of the stations in the borough of Manhattan. This network consists of 166 stations and 256,840 trips. In Figure 10, we show our results for a five-block TDMM-SBM and five-block TDD-SBM for the Manhattan network. The area without stations in the middle of each panel of Figure 10 is Central Park, which has stations on its perimeter but not in its interior.

The blocks that we infer using the five-block TDMM-SBM and five-block TDD-SBM outline similar regions. These models yield block-membership parameters that capture residential and commercial zones much better than the three-block TDMM-SBM and TDD-SBM do for the full

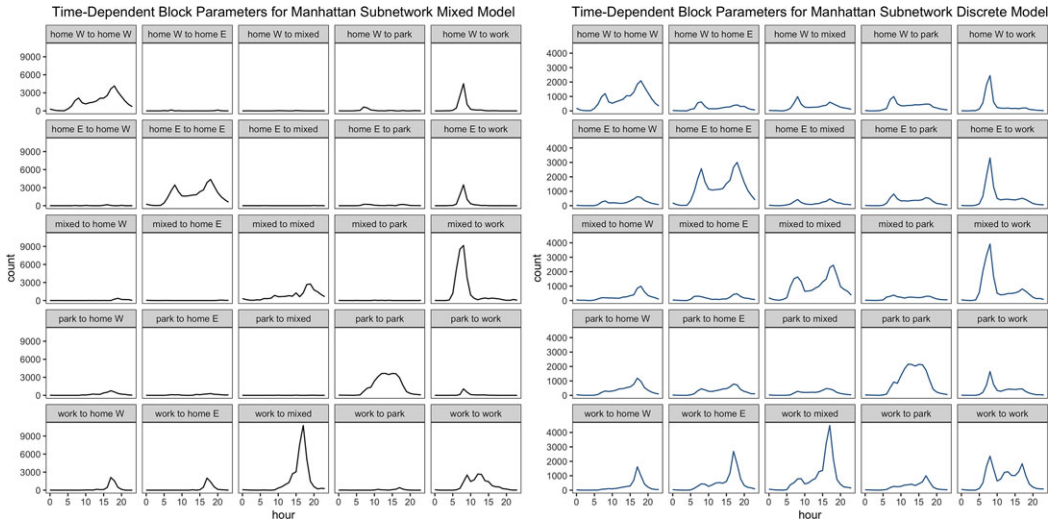


Figure 11. Estimated time-dependent block-connectivity parameters $\hat{\omega}_{ght}$ for (left) the five-block TDMM-SBM and (right) the five-block TDD-SBM for the Manhattan subnetwork of the New York City bicycle-sharing network.

New York City network. One can see this by comparing the five-block results with the underlying zoning map of the area in Figure A.6. The stations in residential zones tend to have larger block-membership parameters for “home” blocks than for “work” blocks, and the opposite is true for stations in commercial zones. We label the five detected blocks as (clockwise from the top left in each panel of Figure 10) “home (west),” “park,” “home (east),” “work,” and “mixed.” We base these labels on the land use of the underlying areas and the time-dependent block-connectivity parameters $\hat{\omega}_{ght}$ (see Figure 11). When we use the TDMM-SBM, the regions tend to blend together; they have sharp boundaries in the TDD-SBM (see Figure 10).

We highlight the appearance of the “park” block, which we have not observed in previous examples and has distinctive usage patterns. The park block is similar to a residential block in its spike in morning traffic to the work block and its spike in evening traffic from the work block, but it has distinctive intra-block activity that peaks in the afternoon. The intra-block activity resembles weekend activity in the New York City bicycle-sharing system as a whole (see Figure 1); this reflects leisure use of the bicycles. Bicycles near Central Park (which also places them near several major museums) are likely to be used by tourists and other non-commuters during the day for leisure or for travel to nearby attractions.

In Figure 11, we show the values of the block-connectivity parameters $\hat{\omega}_{ght}$ for the five-block TDMM-SBM and the five-block TDD-SBM. Our estimates of $\hat{\omega}_{ght}$ for these models illustrate important differences in the usage patterns of different blocks that we observe only when we employ a time-dependent model.⁷ We observe some overlap in the time-dependent usage patterns of the blocks, so there is potential overfitting. For example, the home (east), home (west), and mixed blocks have traffic that is similar to that in blocks other than their own. However, examining the Manhattan network using models with fewer than five blocks does not clearly distinguish the “park” block of stations from other stations in residential zones.

One reason that our time-dependent SBMs for the Manhattan network perform better (with respect to detecting functionally meaningful blocks) than models that we apply to the entire New York City network (see Appendix A.3) is the dependence of station-to-station trip counts on the distances between stations. Although our SBMs correct for the overall activity of each station, they do not normalize expected edge values by the distances between stations. In a small geographical area, such as the coverage areas of the Los Angeles and San Francisco networks, this is a reasonable choice, as all stations are within “biking distance” of each other. However, when

examining a system as large as New York City's, the lack of distance correction weakens the functional roles that we obtain using our time-dependent SBMs. Intra-block trips dwarf inter-block trips (see Figure A.5), and it seems more reasonable to interpret each block as its own ecosystem.

7. Conclusions and discussion

We introduced two types of time-dependent stochastic block models (SBMs) and used them to reveal interesting patterns in urban bicycle-sharing networks. We formulated these SBMs to account both for degree heterogeneity and for a balance between cumulative in-degrees and out-degrees of bicycle-sharing stations over the course of a day. (The latter feature reflects a classical axiom of Ravenstein (1885) that every current of human mobility has an associated countercurrent (Barbosa *et al.*, 2018).) Our SBMs group stations based on their time-dependent usage patterns; this allows us to identify them with home and work roles. Work stations are characterized by inflow from home stations in the morning and outflow to home stations in the afternoon and evening, and residential stations have the opposite flows. It is also sometimes possible to identify other roles, such as a Manhattan park block that combines residential and leisure/touristic behavior.

We illustrated through case studies of Los Angeles, San Francisco, and New York City how our mixed-membership and discrete-membership SBMs can provide complementary insights about transportation patterns in bicycle-sharing systems. We found that many bicycle-sharing stations in our three focal cities have a mixture of roles, which we captured with our mixed-membership SBM. However, in some cases, we observed that a discrete-membership SBM gives a clearer picture of the usage patterns that are associated with each block. We also demonstrated the importance of time-dependence and degree-correction in our models by comparing our results to those from time-independent SBMs and time-dependent SBMs without degree-correction.

We evaluated our block labels by comparing them to city zoning maps. The home–work dichotomies that we detected generally align well with the underlying zones, but there are important deviations near major transit hubs. It is common to evaluate the results of community detection through comparisons with so-called “ground-truth” communities (Fortunato and Hric, 2016; Barbillon *et al.*, 2017) (although it is crucial to be cautious in such evaluations (Peel *et al.*, 2017)). The time-dependent commuting flows that we detected using our SBMs enabled us to identify and label the functional roles of blocks of bicycle-sharing stations even without a corresponding zoning map. In the future, it will be beneficial to compare the activity patterns that we detected to activity patterns from other mobility data, such as taxis, subway systems, e-scooters, and geo-tagged mobile apps (Zhu *et al.*, 2018).

There are several worthwhile ways to improve our models and algorithms. One valuable extension is to allow distributions other than the Poisson distribution for the numbers of trips between stations. Additionally, although it is a convenient simplification to assume independence between the numbers of trips that start in each hour, it may be beneficial to relax this assumption. For example, these numbers are not independent if there are stations that run out of bicycles at some point. As we have discussed, mixed-membership and discrete-membership SBMs can reveal different insights, as can examining SBMs with different numbers of blocks. This helps illustrate why it is important to consider model selection in greater depth. It is also worth considering other estimation techniques for both our mixed-membership and discrete-membership models, especially if one alters the assumption of conditionally independent Poisson edge distributions that we used to simplify our parameter estimation. Variational inference has been employed for many related models (Xing *et al.*, 2010; Yang *et al.*, 2011; Matias and Miele, 2017; Matias *et al.*, 2018), and it is likely that it can also be used in our setting to improve performance (especially for large networks).

Another important direction for future work is the exploration of different methods of pre-processing data to include only the most important edges. The two most apparent ways to do this

are (1) eliminating potentially insignificant edges by thresholding and (2) choosing time layers that reduce variance. The preferential-attachment model of Zhu *et al.* (2013) gives one possible approach for eliminating insignificant edges. The way that one splits the times of a day can improve both accuracy and computational efficiency by reducing the total number of parameters. For example, in the cities that we studied, bicycle trips occurred sporadically between 1 am and 5 am, so it may be desirable to aggregate all of the associated time layers into one time layer to decrease the number of parameters and thereby decrease variance. There exist methods that aim to find suitable ways to segment time periods (Caceres and Berger-Wolf, 2013), and trying to determine the best ones to use in different situations is an active area of research.

Broadening our models to control explicitly for spatial relationships (Barthelemy, 2018) is another natural direction to build on our work. For example, it may enable one to identify functional blocks in geographically diffuse networks such the New York City bicycle-sharing network. The radiation, intervening-opportunities, and gravity models have had some success at describing human mobility over various distances (Barbosa *et al.*, 2018). These mobility models put more weight on longer trips, and some mobility models account for opportunities that lie between origin and destination locations. Some mobility models also possess statistical justification based on entropy arguments, and it is worthwhile to investigate methods to incorporate them into SBMs. Mobility models were incorporated into null models in time-dependent modularity objective functions in Sarzynska *et al.* (2016), who found that radiation and gravity null models are more appropriate than the usual Newman–Girvan null model (which is a variant of a configuration model (Fosdick *et al.*, 2018)) for spatial networks. One can use similar ideas to incorporate spatial information into SBMs. The value of using spatial null models in studies of bicycle-sharing systems has been examined previously (Austwick *et al.*, 2013; Taiyeb, 2014), so this is a very interesting direction to pursue.

It is also worthwhile to use our time time-dependent SBMs and extensions of them to study other types of mobility data. One example is dockless vehicle-sharing networks, such as those from e-scooter-sharing services. To make the data from such systems amenable to analysis using our models, one can partition a city into a grid and measure the usage in each region, taking care to recognize irregularities from transit hubs. Depending on how heavily these systems are used in commuting, one may discover functional blocks other than ones for “home” and “work.” We expect that it will also be possible to tailor our methods to analyze multimodal transportation systems and other urban flows, which are particularly suitable for analysis using multilayer networks (Kivelä *et al.*, 2014; Aleta and Moreno, 2019; Natera Orozco, 2020; Gallotti *et al.*, 2021).

Acknowledgments. We thank Brian Karrer and Mark Newman for allowing us to use and share their code for degree-corrected SBMs from Karrer and Newman (2011). We thank Susan Handy’s lab at UC Davis for useful discussions on contextualizing our work for transportation researchers and planners, David Kempe for pointing us to Caceres and Berger-Wolf (2013), and Michelle Feng and others in the networks journal club at UCLA for helpful comments. We also thank two anonymous referees for many helpful comments and Phil Chodrow for technical assistance with setting up computations. CM and SSC thank NSF (DMS-1351860) for funding, and SSC also thanks NIGMS (R01 GM126556) and an NIH Ruth L. Kirschstein National Research Service Award (T32-GM008185) for funding. SW thanks NSF (CCF-1422795), ONR (N000141110719, N000141210834), DOD (W81XWH-15-1-0147), Intel STC-Visual Computing Grant (20112360), and Adobe Inc. for funding.

Competing interests. None.

Notes

1 We take extremely short trips to be those that last 2 minutes or less. We take extremely long trips to be those that last 90 minutes or more in Los Angeles and San Francisco and 120 minutes or more in New York City (because of the larger coverage areas of stations in New York City).

2 In the bicycle-sharing networks that motivate our work, interlayer edges (which start in one time layer and end in another) can help capture trip lengths, which are already captured in the inter-station distances that are associated with the edges in our

networks. Including interlayer edges is unlikely to add to the insights that we gain from our models, and it would significantly increase their complexity.

3 When describing weighted networks, we use the term “degree” to refer to the weighted degree. We use the terms “in-degree” and “out-degree” to describe the weighted in-degree and weighted out-degree, respectively.

4 Although we are maximizing a function and thus technically performing gradient ascent, we refer to this method by its more common monicker of “gradient descent.”

5 We also apply our TDD-SBM to the London bicycle-sharing network that was studied by Matias *et al.* (2018) in the code at https://github.com/jcarlen/tdsbm_supplementary_material.

6 Permission for use of these proprietary data is granted by the City of Los Angeles Department of City Planning. Copyright © 2015 City of Los Angeles. All Rights Reserved.

7 We obtain similar blocks for this network using a discrete, directed, degree-corrected, time-independent SBM as we do using a TDD-SBM with the same number of blocks. We do not show our results from this time-independent SBM, but one can produce them using the code that we provide at https://github.com/jcarlen/tdsbm_supplementary_material.

References

- Airoldi, E. M., Blei, D. M., Fienberg, S. E., & Xing, E. P. (2008). Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9, 1981–2014.
- Akbarzadeh, M., Mohri, S. S., & Yazdian, E. (2018). Designing bike networks using the concept of network clusters. *Applied Network Science*, 3(1), 12.
- Akaike, H. (1974). A new look at the statistical model identification. In *Selected Papers of Hirotugu Akaike* (pp. 215–222). Heidelberg, Germany: Springer-Verlag.
- Aleta, A., & Moreno, Y. (2019). Multilayer networks in a nutshell. *Annual Review of Condensed Matter Physics*, 10, 45–62.
- Ashqar, H. I., Elhenawy, M., & Rakha, H. A. (2019). Modeling bike counts in a bike-sharing system considering the effect of weather conditions. *Case Studies on Transport Policy*, 7(2), 261–268.
- Ashqar, H. I., Elhenawy, M., Rakha, H. A., Almannaa, M., & House, L. (2021). Network and station-level bike-sharing system prediction: A San Francisco Bay Area case study. *Journal of Intelligent Transportation Systems*, doi: 10.1080/15472450.2021.1948412.
- Austwick, M. Z., O’Brien, O., Strano, E., & Viana, M. (2013). The structure of spatial networks and communities in bicycle sharing systems. *PLoS One*, 8(9), e74685.
- Barbillon, P., Donnet, S., Lazega, E., & Bar-Hen, A. (2017). Stochastic block models for multiplex networks: An application to a multilevel network of researchers. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(1), 295–314.
- Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J. J., Simini, F., & Tomasini, M. (2018). Human mobility: Models and applications. *Physics Reports*, 734, 1–74.
- Barthelemy, M. (2018). *Morphogenesis of Spatial Networks*. Cham, Switzerland: Springer International Publishing.
- Barthelemy, M. (2019). The statistical physics of cities. *Nature Reviews Physics*, 1, 406–415.
- Bay Area Bike Share. (2017). Bay Area Bike Share Open Data. Available at https://s3.amazonaws.com/babs-open-data/babs_open_data_year_3.zip; last checked 2019-05-05; archived main site: <https://web.archive.org/web/20170303021745/http://www.bayareabikeshare.com/open-data>.
- Bazzi, M., Jeub, L. G. S., Arenas, A., Howison, S. D., & Porter, M. A. (2020). A framework for the construction of generative models for mesoscale structure in multilayer networks. *Physical Review Research*, 2, 023100.
- Bike share in the U.S.: 2017. (May 2018). Tech. rep., National Association of City Transportation Officials. Available at <https://nacto.org/bike-share-statistics-2017/>.
- Borgnat, P., Abry, P., Flandrin, P., Robardet, C., Rouquier, J.-B., & Fleury, E. (2011). Shared bicycles in a city: A signal processing and data analysis perspective. *Advances in Complex Systems*, 14(03), 415–438.
- Caceres, R. S., & Berger-Wolf, T. (2013). Temporal scale of dynamic networks. In P. Holme & J. Saramäki (Eds). *Temporal Networks* (pp. 65–94) Heidelberg, Germany: Springer-Verlag.
- Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 1–32.
- Cranmer, S. J., Menninga, E. J., & Mucha, P. J. (2015). Kantian fractionalization predicts the conflict propensity of the international system. *Proceedings of the National Academy of Sciences of the United States of America*, 112(38), 11812–11816.
- Department of City Planning, City of Los Angeles (2022). Regulations summary. Available at <https://planning.lacity.org/zoning/regulations-summary> (accessed 10 February 2022).
- Department of City Planning, City of Los Angeles (2015). City of Los Angeles GIS data: Zoning. Available at <http://planning.lacity.org/>.
- Department of City Planning, New York City (2018). NYC zoning districts. Available at <https://www1.nyc.gov/site/planning/data-maps/open-data/dwn-gis-zoning.page>.
- Edelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8), 1–18.

- Etienne, C., & Latifa, O. (2014). Model-based count series clustering for bike sharing system usage mining: A case study with the Vélib' system of Paris. *ACM Transactions on Intelligent Systems and Technology*, 5(3), 39.
- Faust, K., & Wasserman, S. (1992). Blockmodels: Interpretation and evaluation. *Social Networks*, 14(1–2), 5–61.
- Fishman, E., Washington, S., & Haworth, N. (2013). Bike share: A synthesis of the literature. *Transport Reviews*, 33(2), 148–165.
- Forma, I., Raviv, T., & Tzur, M. (2015). A 3-step math heuristic for the static repositioning problem in bike-sharing systems. *Transportation Research Part B: Methodological*, 71, 230–247.
- Fortunato, S., & Hric, D. (2016). Community detection in networks: A user guide. *Physics Reports*, 659, 1–44.
- Fosdick, B. K., Larremore, D. B., Nishimura, J., & Ugander, J. (2018). Configuring random graph models with fixed degree sequences. *SIAM Review*, 60(2), 315–355.
- Funke, T., & Becker, T. (2019). Stochastic block models: A comparison of variants and inference methods. *PLoS ONE*, 14(4), e0215296.
- Gallotti, R., Bertagnolli, G., & De Domenico, M. (2021). Unraveling the hidden organisation of urban systems and their mobility flows. *The European Physical Journal—Data Science*, 10, 3.
- Giorgi, D., Matias, C., Rebafka, T., & Villers, F. (2018). *PPSBM: Clustering in Longitudinal Networks*. R package version 0.2.2. Available at <https://CRAN.R-project.org/package=ppsbm>.
- Griffin, G. P., & Sener, I. N. (2016). Planning for bike share connectivity to rail transit. *Journal of Public Transportation*, 19(2), 1–22.
- He, M., Glasser, J., Bhamidi, S., & Kaza, N. (2019). Intertemporal community detection in bikeshare networks. arXiv:1906.04582.
- Holland, P. W., Laskey, K. B., & Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks*, 5(2), 109–137.
- Holme, P., & Saramäki, J. (2012). Temporal networks. *Physics Reports*, 519(3), 97–125.
- Holme, P. (2015). Modern temporal network theory: A colloquium. *The European Physical Journal B*, 88, 234.
- Ho, Q., Song, L., & Xing, E. P. (2011). Evolving cluster mixed-membership blockmodel for time-varying networks. In G. Gordon, D. Dunson, & M. Dudik (Eds.) *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics* (pp. 342–350).
- Hu, J., Qin, H., Yan, T., & Zhao, Y. (2020). Corrected Bayesian information criterion for stochastic block models. *Journal of the American Statistical Association*, 115(532), 1771–1783.
- Jeub, L. G. S., Mahoney, M. W., Mucha, P. J., & Porter, M. A. (2017). A local perspective on community structure in multilayer networks. *Network Science*, 5(2), 144–163.
- Karrer, B., & Newman, M. E. J. (2011). Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1), 016107.
- Kernighan, B. W., & Lin, S. (1970). An efficient heuristic procedure for partitioning graphs. *Bell System Technical Journal*, 49(2), 291–307.
- Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., & Porter, M. A. (2014). Multilayer networks. *Journal of Complex Networks*, 2(3), 203–271.
- Kobayashi, T., Takaguchi, T., & Barrat, A. (2019). The structured backbone of temporal social ties. *Nature Communications*, 10(1), 220.
- LA Metro (2017). Metro Bike Share Trip Data. Available at https://bikeshare.metro.net/wp-content/uploads/2017/01/Metro_trips_Q4_2016.zip; last checked 2019-05-05; the version was last modified on 2018-09-17.
- Latouche, P., Birmelé, E., & Ambroise, C. (2011). Overlapping stochastic block models with application to the French political blogosphere. *The Annals of Applied Statistics*, 5(1), 309–336.
- Leger, J.-B., Barbillon, P., & Chiquet, J. (2020). *Blockmodels: Latent and Stochastic Block Model Estimation by a 'V-EM' Algorithm*. R package version 1.1.4. Available at <https://cran.r-project.org/web/packages/blockmodels/index.html>.
- Lorrain, F., & White, H. (1971). Structural equivalence of individuals in social networks. *Journal of Mathematical Sociology*, 1(1), 49–80.
- Matias, C., & Miele, V. (2017). Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4), 1119–1141.
- Matias, C., Rebafka, T., & Villers, F. (2018). A semiparametric extension of the stochastic block model for longitudinal networks. *Biometrika*, 105, 665–680.
- Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., & Onnela, J.-P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 328(5980), 876–878.
- Munoz-Mendez, F., Han, K., Klemmer, K., & Jarvis, S. (2018). Community structures, interactions and dynamics in London's bicycle sharing network. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, UbiComp '18* (pp. 1015–1023). New York, NY, USA: Association for Computing Machinery.
- Natera Orozco, L. G., Battiston, F., Iñiguez, G., & Szell, M. (2020). Data-driven strategies for optimal bicycle network growth. *Royal Society Open Science*, 7, 201130.
- Newman, M. E. J. (2018). *Networks* (2nd ed.). Oxford, UK: Oxford University Press.
- Nowicki, K., & Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455), 1077–1087.

- NYCBS, Citi Bike System Data. Available at <https://s3.amazonaws.com/tripdata/201610-citibike-tripdata.zip>; last checked 2019-05-05.
- Papadopoulos, L., Porter, M. A., Daniels, K. E., & Bassett, D. S. (2018). Network analysis of particles and grains. *Journal of Complex Networks*, 6(4), 485–565.
- Paul, S., & Chen, Y. (2021). Null models and community detection in multi-layer networks. *Sankhyā A*. <https://doi.org/10.1007/s13171-021-00257-0>.
- Peel, L., Larremore, D. B., & Clauset, A. (2017). The ground truth about metadata and community detection in networks. *Science Advances*, 3(5), e1602548.
- Peixoto, T. P. (2019). Bayesian Stochastic Blockmodeling. In P. Doreian, V. Batagelj, & A. Ferlijug (Eds.), *Advances in Network Clustering and Blockmodeling*, ch. 11 (pp. 289–332). Hoboken, NJ, USA: John Wiley & Sons, Ltd.
- Pfommer, J., Warrington, J., Schildbach, G., & Morari, M. (2014). Dynamic vehicle redistribution and online price incentives in shared mobility systems. *IEEE Transactions on Intelligent Transportation Systems*, 15, 1567–1578.
- Porter, M. A. (2018). WHAT IS... A multilayer network. *Notices of the American Mathematical Society*, 65(11), 1419–1423.
- Ravenstein, E. J. (1885). The laws of migration. *Journal of the Royal Statistical Society of London*, 48(2), 167–235.
- R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Romanillos, G., & Zaltz Austwick, M. (2016). Madrid cycle track: Visualizing the cyclable city. *Journal of Maps*, 12(5), 1218–1226.
- Romanillos, G., Moya-Gómez, B., Zaltz-Austwick, M., & Lamíquiz-Daudén, P. J. (2018). The pulse of the cycling city: Visualising Madrid bike share system GPS routes and cycling flow. *Journal of Maps*, 14(1), 34–43.
- Rossi, R. A., & Ahmed, N. K. (2015). Role discovery in networks. *IEEE Transactions on Knowledge and Data Engineering*, 27, 1112–1131.
- Rossetti, G., & Cazabet, R. (2018). Community discovery in dynamic networks: A survey. *ACM Computing Surveys*, 51(2), 35.
- Rosvall, M., Axelsson, D., & Bergstrom, C. T. (2009). The map equation. *The European Physical Journal Special Topics*, 178(1), 13–23.
- Saldaña, D. F., Yu, Y., & Feng, Y. (2017). How many communities are there? *Journal of Computational and Graphical Statistics*, 26(1), 171–181.
- Sarzynska, M., Leicht, E. A., Chowell, G., & Porter, M. A. (2016). Null models for community detection in spatially embedded, temporal networks. *Journal of Complex Networks*, 4(3), 363–406.
- Schmidt, C. (2018). Active travel for all? The surge in public bike-sharing programs. *Environmental Health Perspectives*, 126(8), 082001.
- Shu, J., Chou, M., Liu, Q., Teo, C., & Wang, I.-L. (2013). Models for effective deployment and redistribution of bicycles within public bicycle-sharing systems. *Operations Research*, 61(6), 1346–1359.
- Singhvi, D., Singhvi, S., Frazier, P. I., Henderson, S. G., Mahony, E. O., Shmoys, D. B., & Woodard, D. B. (2015). Predicting bike usage for New York City's bike sharing system. In *Proceedings of the Association for the Advancement of Artificial Intelligence*.
- Snijders, T. A. B., & Nowicki, K. (1997). Estimation and prediction for stochastic block-structures for graphs with latent block structure. *Journal of Classification*, 14, 75–100.
- Stan Development Team. (2018). RStan: The R interface to Stan. R package version 2.18.2. Available at <https://cran.r-project.org/web/packages/rstan/vignettes/rstan.html>.
- Stanley, N., Shai, S., Taylor, D., & Mucha, P. J. (2016). Clustering network layers with the strata multilayer stochastic block model. *IEEE Transactions on Network Science and Engineering*, 3(2), 95–105.
- Taiyeb, A. (2014). TP19: Spatial networks and human mobility: An application of the intervening opportunities model to the London cycle hire scheme. Undergraduate Thesis, Department of Physics, University of Oxford. Available at <https://www.math.ucla.edu/mason/research/aqil-TP19-final.pdf>.
- Valdano, E., Ferreri, L., Poletto, C., & Colizza, V. (2015). Analytical computation of the epidemic threshold on temporal networks. *Physical Review X*, 5(2), 021005.
- Valles-Catala, T., Massucci, F. A., Guimera, R., & Sales-Pardo, M. (2016). Multilayer stochastic block models reveal the multilayer structure of complex networks. *Physical Review X*, 6(1), 011036.
- Vavrek, M. J. (2011). *fossil*: Palaeoecological and palaeogeographical analysis tools. *Palaeontologia Electronica*, 14(1), 1T. R package version 0.4.0. Available at https://palaeo-electronica.org/2011_1/238/index.html.
- Wang, Y. X. R., & Bickel, P. J. (2017). Likelihood-based model selection for stochastic block models. *The Annals of Statistics*, 45(2), 500–528.
- Weikel, D., & Karlamangla, S. (2015). San Francisco residents relying less on private automobiles. *Los Angeles Times*, February 23.
- Wergin, J., & Buehler, R. (2017). Where do bikeshare bikes actually go?: Analysis of capital bikeshare trips with GPS data. *Transportation Research Record*, 2662(1), 12–21.
- Xing, E. P., Fu, W., & Song, L. (2010). A state-space mixed membership blockmodel for dynamic network tomography. *The Annals of Applied Statistics*, 4(2), 535–566.
- Xie, X.-F., & Wang, Z. J. (2018). Examining travel patterns and characteristics in a bikesharing network and implications for data-driven decision supports: Case study in the Washington DC area. *Journal of Transport Geography*, 71, 84–102.

Xu, K. S., & Hero, A. O. (2014). Dynamic stochastic blockmodels for time-evolving social networks. *IEEE Journal of Selected Topics in Signal Processing*, 8(4), 552–562.

Yan, X. (2016). Bayesian model selection of stochastic block models. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM '16* (pp. 323–328). Piscataway, NJ, USA: IEEE Press.

Yan, X., Shalizi, C., Jensen, J. E., Krzakala, F., Moore, C., Zdeborová, L., Zhang, P., & Zhu, Y. (2014). Model selection for degree-corrected block models. *Journal of Statistical Mechanics: Theory and Experiment*, 2014(5), P05007.

Yang, T., Chi, Y., Zhu, S., Gong, Y., & Jin, R. (2011). Detecting communities and their evolutions in dynamic social networks—A Bayesian approach. *Machine Learning*, 82(2), 157–189.

Zhang, X., Moore, C., & Newman, M. E. J. (2017). Random graph models for dynamic networks. *The European Physical Journal B*, 90(10), 200.

Zhu, E., Khan, M., Kats, P., Santosh Bamne, S., & Sobolevsky, S. (2018). Digital urban sensing: A multi-layered approach. arXiv:1809.01280.

Zhu, Y., Yan, X., & Moore, C. (2013). Oriented and degree-generated block models: Generating and inferring communities with inhomogeneous degree distributions. *Journal of Complex Networks*, 2(1), 1–18.

Appendix

A.1 Singular vectors for bicycle-sharing data

To inform our modeling approach, we explore the imbalances between morning and evening activity in the stations of each network. We calculate singular-value decompositions (SVDs) of the matrices of the hourly in-degree and out-degree of each station. To be explicit, entry i, j of the matrix of in-degrees is equal to the total number of trips that arrive at station i in hour j , and we construct the matrix of out-degrees analogously for departing trips. We show results for Los Angeles, San Francisco, and New York City in Figures A.1, A.2, and A.3, respectively. The first two principle components either strengthen both observed peaks or weaken one peak while strengthening the other. Across the three data sets, the first two singular vectors explain between 88% and 97% of the variation in the corresponding matrix, supporting the importance of peak morning and peak evening commutes for classifying stations.

A.2 Proof that the expected node degrees of the MLE of the TDD-SBM equal the observed node degrees

Suppose that we generate a network from the MLE of our time-dependent discrete-membership stochastic block model (TDD-SBM). We prove that the expected value of the degree of a node (i.e., the sum of its in-degree and its out-degree over all time layers) is the same as the degree of the node in the observed data. That is, we prove that $\sum_j \sum_{t=0}^{23} (\mu_{ijt} + \mu_{jit}) = k_i = \sum_j \sum_{t=0}^{23} A_{ijt}$.

Let i and j be nodes of the network, and let $t \in \{0, 1, \dots, 23\}$ index the time layers. Additionally, recall that g denotes block g in the TDD-SBM and that κ_g is the sum of the in-degrees and the out-degrees of all nodes in block g over all time layers. We calculate

$$\begin{aligned} \sum_j \sum_{t=0}^{23} (\mu_{ijt} + \mu_{jit}) &= \sum_j \sum_{t=0}^{23} \theta_i \theta_j (\omega_{g_i g_j t} + \omega_{g_j g_i t}) \\ &= \frac{k_i}{\kappa_{g_i}} \sum_t \sum_h \sum_{j \in h} \frac{k_j}{\kappa_h} (m_{g_i h t} + m_{h g_i t}) \\ &= \frac{k_i}{\kappa_{g_i}} \sum_t \sum_h (m_{g_i h t} + m_{h g_i t}) \\ &= k_i . \end{aligned}$$

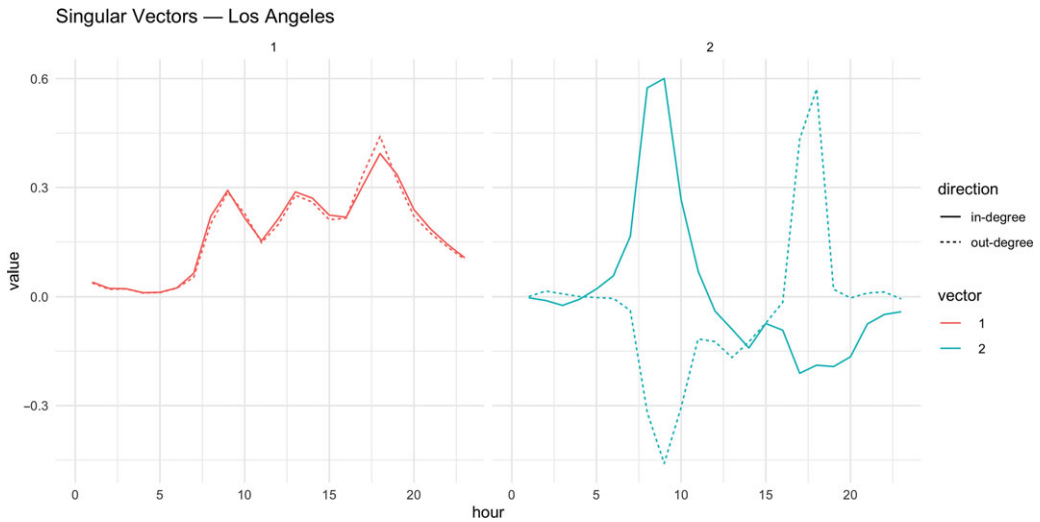


Figure A.1. The first two singular vectors of the in-degree matrix and the out-degree matrix of the downtown Los Angeles bicycle-sharing network.

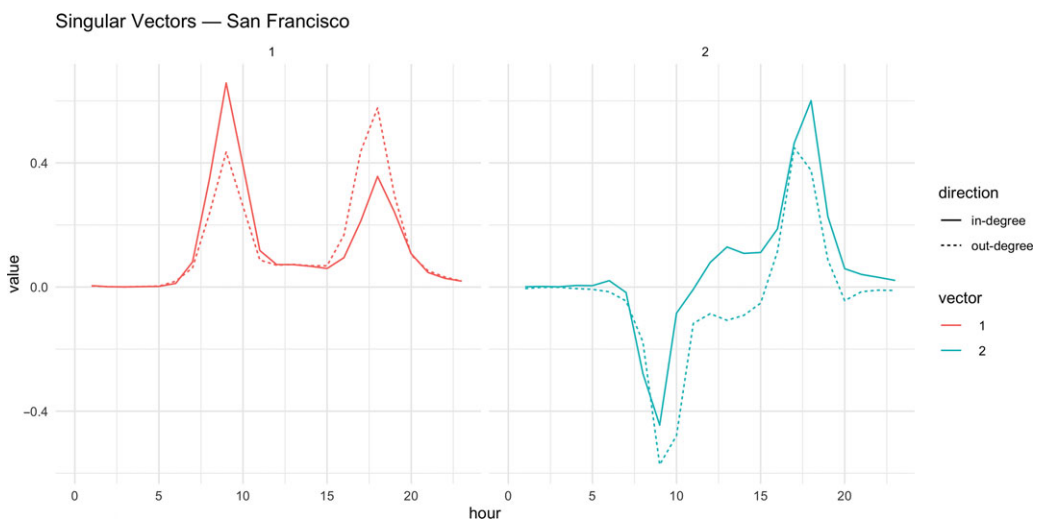


Figure A.2. The first two singular vectors of the in-degree matrix and the out-degree matrix of the San Francisco bicycle-sharing network.

A.3 TDMM-SBM and TDD-SBM for the entire New York City bicycle-sharing network

In Figure A.4, we compare our results from a three-block TDMM-SBM and a three-block TDD-SBM for the New York City bicycle-sharing network. In initial calculations, we found that a two-block TDD-SBM divides the network along the East River into a Manhattan block and Brooklyn block and that a two-block TDMM-SBM divides the network slightly farther north in Lower Manhattan. We do not show the outputs of the two-block models, but we provide the replication code, estimated parameters, and associated visualizations at https://github.com/jcarlen/tdsbm_supplementary_material.

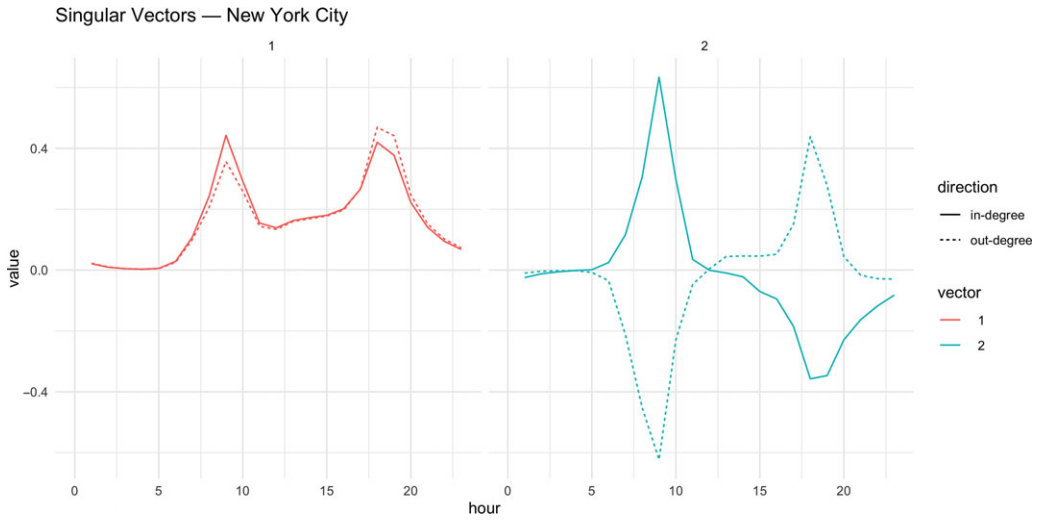


Figure A.3. The first two singular vectors of the in-degree matrix and the out-degree matrix of the New York City bicycle-sharing network.

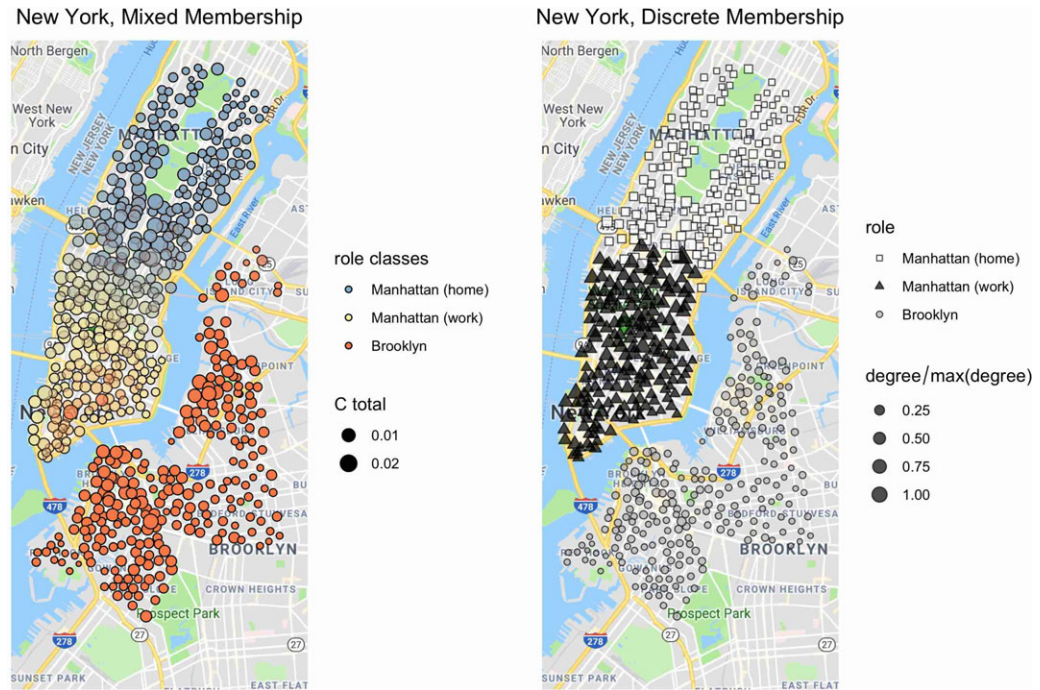


Figure A.4. Classification of New York City bicycle-sharing stations using (left) a three-block TDDM-SBM and (right) a three-block TDD-SBM. The sizes of the nodes take continuous values. In the left panel, we scale the area of each node based on the value of $\sum_g C_{ig}$. In the right panel, we scale the area of each node based on its degree.

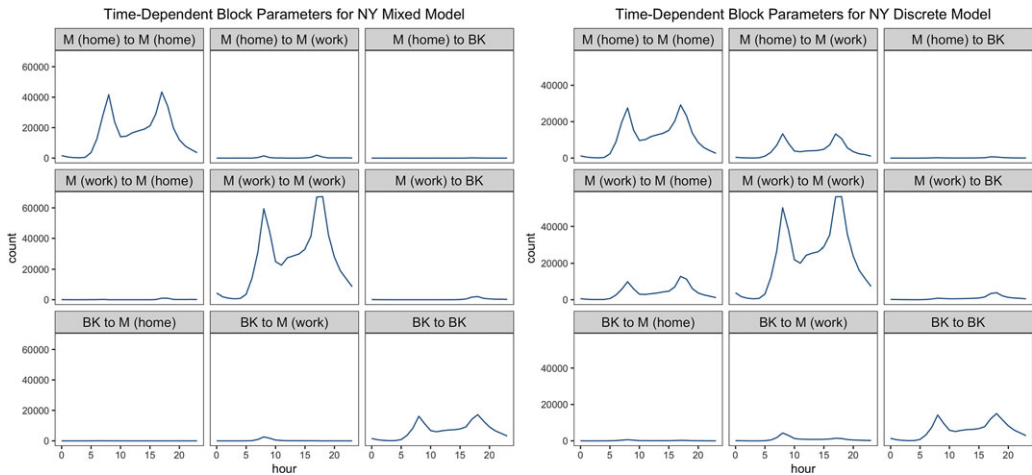


Figure A.5. Estimated time-dependent block-connectivity parameters $\hat{\omega}_{ght}$ for (left) the three-block TDMM-SBM and (right) the three-block TDD-SBM for the New York City bicycle-sharing network. We use “M” to signify Manhattan and “BK” to signify Brooklyn.

In Figure A.5, we compare the inferred inter-block traffic, as captured by the values of $\hat{\omega}_{ght}$, for the three-block TDMM-SBM and three-block TDD-SBM. We observe prominently that all intra-block traffic has two peaks and much higher hourly trip counts than inter-block traffic. The double peaks are reminiscent of the overall system activity in Figure 1. This may be due in part to last-mile commuting, as we saw in the San Francisco bicycle-sharing network. However, for a system that is this geographically large, the two peaks and minimal inter-block traffic suggest that it is useful (and important) to consider each block as its own ecosystem. There is also a strong similarity between the results from our TDD-SBM and the results (not shown) from a three-block time-independent SBM for a time-aggregated New York City network, providing further evidence that our time-dependent SBMs are not capturing time-dependent roles for New York City when we consider its entire bicycle-sharing system. Consequently, we choose the labels of these blocks (see Figures A.4 and A.5) based on the primary borough and zone type of each block’s stations, as indicated in the zoning map for this part of New York City (Department of City Planning, New York City, 2018) in Figure A.6.

In Figure A.6, we illustrate that there is general overlap, although it is far from perfect, between (1) the Upper Manhattan (“home”) block and residential areas or parks and between (2) the Lower Manhattan (“work”) block and commercial or manufacturing areas. All stations in Brooklyn are in the third block, which mostly has residential areas. We show only the blocks that we infer using our TDD-SBM in Figure A.6, but the same reasoning motivates our labels for the three-block TDMM-SBM. No block has exclusively commercial or residential areas.

We examined several time-dependent SBMs for New York City with more than three blocks to try to discover functional blocks, but we found that the blocks were still primarily geographical. In some cases, the TDMM-SBM with four or more blocks yielded functional divisions within smaller geographical areas (subdividing the blocks in Figure A.4), but none of our models detected system-wide “home” or “work” blocks. See our code and figures at https://github.com/jcarlen/tdsbm_supplementary_material to generate and view the results of applying time-dependent SBMs with 2–6 blocks to the New York City bicycle-sharing network.

Across our models, we find that the distance between two stations in the New York City bicycle-sharing network is the feature that is associated most strongly with whether or not those stations belong to the same block. As we noted above, the detected blocks represent primarily geographical divisions, with most of the traffic occurring within blocks. This contrasts with our results for the Los Angeles and San Francisco networks. We believe that this occurs because the New York City

Discrete Roles versus NYC Zoning

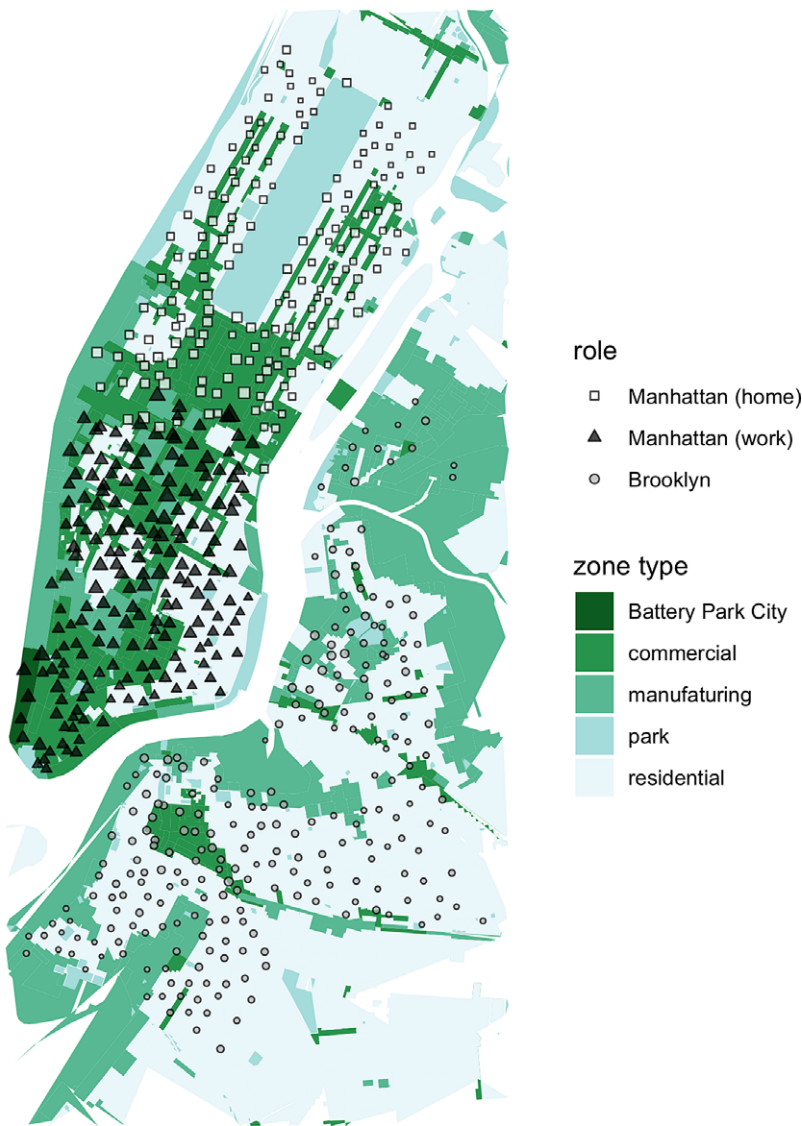


Figure A.6. Three-block TDD-SBM station roles versus the coverage-area zoning map of New York City.

bicycle-sharing network is geographically larger than the San Francisco and Los Angeles bicycle-sharing networks. Our models are useful for detecting functional blocks only when the distances between stations are not the main determinant of inter-station traffic over time.

A.4 Model selection

Although statistically rigorous model selection is outside the scope of our paper, we briefly compare the number of parameters in our mixed-membership SBM (TDMM-SBM) and discrete-membership SBM (TDD-SBM). This is valuable for considering model-selection criteria, such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC),

Table A.1. Comparison of log-likelihoods and numbers of parameters (N_p) in SBMs for the Manhattan network, which has $N = 166$ nodes

Blocks	TDMM-SBM		TDD-SBM	
	N_p	log-likelihood	N_p	log-likelihood
2	426	-260,625	426	-270,809
3	711	-235,162	545	-254,779
4	1,044	-212,295	712	-236,198
5	1,425	-198,489	927	-222,539
6	1,854	-189,670	1,190	-216,468

that penalize a model based on its number of parameters. For a network with N nodes, K blocks, and T time layers, the number of parameters in the TDMM-SBM is

$$KN - K + TK^2 \quad (8)$$

and the number of parameters in the TDD-SBM is

$$2N - K + TK^2. \quad (9)$$

The first term of (8) comes from the fact that each node in the TDMM-SBM has K parameters (C_{ig} , with $g \in \{1, \dots, K\}$) that express the strength of membership in each block. By contrast, each node in the TDD-SBM has one parameter for block membership and one degree-correction parameter. Therefore, given a value of N , the first term in (8) increases linearly with the number of blocks, whereas the corresponding term in (9) is fixed. Otherwise, formulas (8) and (9) are identical. The $-K$ term in each formula arises from the identifiability constraints in each model. As we described in Section 3, these constraints are $\sum_i C_{ig} = 1$ (for each block g) for the TDMM-SBM and $\sum_{i \in g} \theta_i = 1$ (for each block g) for TDD-SBM. The last term (i.e., TK^2) in each formula is the total number of ω_{ght} terms in the model (see Section 3.1).

We show the unnormalized log-likelihoods and numbers of parameters (N_p) in the TDMM-SBM and TDD-SBM for the Manhattan (home) subnetwork (which has $N = 166$ nodes) with 2–6 blocks (see Table A.1). In this example, the TDMM-SBM outperforms the TDD-SBM with respect to log-likelihood when the two models have the same number of parameters. This result makes sense because of the additional constraint of the TDD-SBM that each station must belong to exactly one block.

In Figure A.7, we plot the Akaike information criterion (AIC) (Akaike, 1974), which is given by

$$\text{AIC} = (2 \times N_p) - (2 \times \text{log-likelihood}),$$

versus the number of blocks in our TDMM-SBM for the Los Angeles bicycle-sharing network with 2–10 blocks. We consider models with a larger numbers of blocks in the Los Angeles network than we did in the Manhattan network because the former's smaller number of stations allows faster computations.

The AIC is a cost function for comparing the relative qualities of statistical models. Given two models, one regards the model with a smaller AIC as the “better” model. The AIC accounts for both the likelihood and the complexity of a model, where one measures the latter based on the number of parameters in the model. In Figure A.7, we see that the AIC of the TDMM-SBM for the Los Angeles bicycle-sharing network decreases as we increase the number of blocks from 2 to 10. Based on the AIC criterion, one would select the 10-block model. However, plots of the block-connectivity parameters $\hat{\omega}_{ght}$ for the TDMM-SBM for the Los Angeles network with 7 or more blocks are noisier and no more informative than ones from the TDMM-SBM with fewer blocks. These plots suggest strongly that there is overfitting in these cases. This is in line with experiments

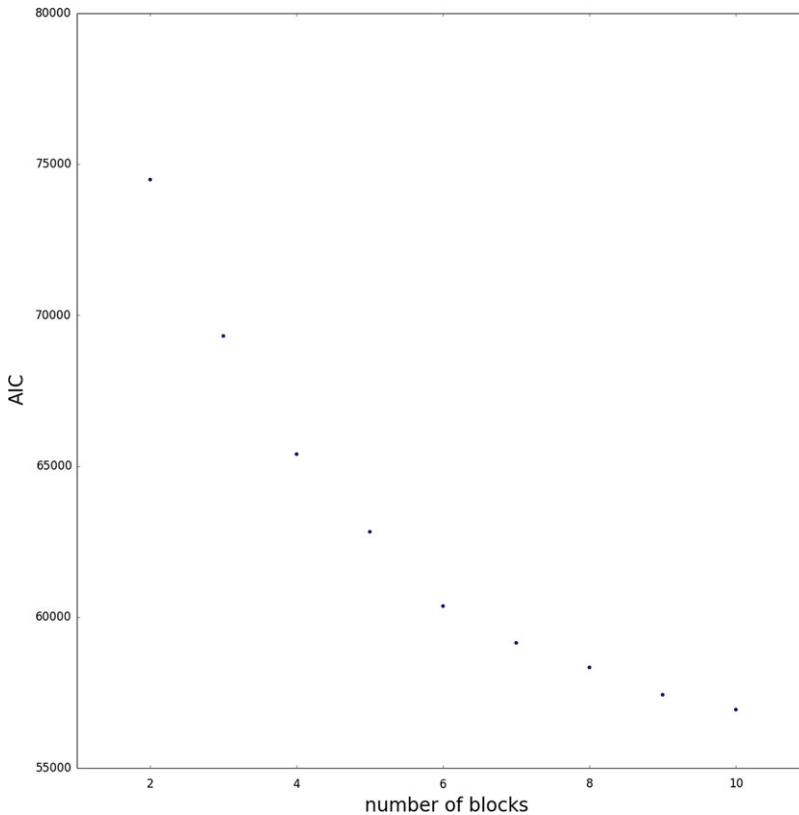


Figure A.7. Values of the Akaike information criterion for the MLE of the TDMM-SBM with 2–10 blocks for the Los Angeles bicycle-sharing network.

by Funke and Becker (2019), who found that the AIC overestimates the number of blocks for several types of SBMs.

Our time-dependent SBMs are incomplete descriptions of the data-generating process in bicycle-sharing networks. For example, they do not account for the distances between stations. The AIC tends to recommend a model with more blocks to compensate for this incompleteness, but (as we just discussed) visual inspection reveals the likely overfitting that results from including too many blocks. We seek to gain insight into bicycle-sharing systems, and we prefer models with fewer blocks that describe station roles in broad but interpretable terms over less-interpretable models with more blocks but smaller AIC values. Choosing appropriate model-selection criteria for SBMs deserves additional consideration and is an active research area (Yan *et al.*, 2014; Yan, 2016; Saldaña *et al.*, 2017; Wang and Bickel, 2017; Hu *et al.*, 2020). We leave such investigations for future work.