Equivalence and Clustering in Worker Flows:

Stochastic Blockmodels for the Analysis of Mobility Tables

Barum Park*

Abstract

Mobility scholars are increasingly turning to computational methods to analyze mobility tables. Most of these approaches start with the detection of mobility clusters: namely, sets of occupations within which the flow of workers is dense and across which it is sparse. Yet, clustering is not the only way in which worker flows can be structured. This paper shows how a degree-corrected stochastic blockmodel is able to detect patterns of mobility that are more general than clustering and consistent with the homogeneity criterion laid out by Goodman (1981) as well as the internal homogeneity thesis proposed by Breiger (1981). Due to the intractable marginal likelihood of the model, parameters are estimated via a variational Expectation Maximization algorithm. Simulation results suggest that the estimation algorithm successfully recovers (conditionally) stochastically equivalent mobility classes Further, the analysis of two real-world examples shows that the model is able to detect meaningful mobility patterns, even in situations where commonly used community detection algorithms fail.

Keywords: Mobility Table Analysis, Homogeneity, Internal Homogeneity Thesis, Stochastic Equivalence, Stochastic Blockmodels, Variational EM

*This research was supported by the National Science Foundation (Award no. 2116938) and the Cornell Center for Social Science Faculty Fellows program. I thank Siwei Cheng, Michael Hout, BK Lee, and the three anonymous reviewers for their invaluable comments. Computational resources provided by the CCSS Cloud Computing Solutions are gratefully acknowledged. An earlier version of this work has been presented at the Annual meeting of the American Sociological Association, August 2022, Los Angeles, where I received critical and productive feedback from Ron Breiger and Yu Xie. All remaining errors are my own. Direct correspondence to Barum Park, Assistant Professor, Department of Sociology, Cornell University (338 Uris Hall, Cornell University, Ithaca, NY 14853. Email: b.park@cornell.edu). Replication materials are available at `https://osf.io/enhdw/`, which includes an `R` package that was used to fit the degree-corrected stochastic blockmodels presented in this paper.

Introduction

Sociologists are increasingly turning to computational methods to detect structure in mobility tables. Many of these new approaches treat the mobility table as a weighted network, where occupations are represented as nodes and the amount of worker flow between them as weighted edges. This reconceptualization has enabled researchers to apply recent developments in clustering and community detection algorithms to mobility tables, pushing forward the boundaries of mobility research (Schmutte 2014; Melamed 2015; Toubøl and Larsen 2017; Cheng and Park 2020; Lin and Hung 2022). Notwithstanding their merits, however, the exclusive focus on within- versus between-cluster density of these approaches overlooks alternative ways in which mobility can be structured (but see, Block et al. 2022). Indeed, clustering is agnostic to the *concrete pattern* of worker flow across clusters as long as the intra-cluster density is high enough and, hence, stops halfway before arriving at what network scholars have traditionally regarded as "positions" within the web of connections (White et al. 1976).

In this paper, I use a degree-corrected stochastic blockmodel (DCSBM) as an alternative approach to finding aggregation schemes from mobility tables. The model can be understood as a natural extension of the log-linear model (Hout 1983; Agresti 2003) with discrete latent variables that represent the class membership of occupations. These classes, in turn, represent conditionally "stochastically equivalent" (Holland et al. 1983) positions, meaning that occupations belonging to the same class share the same expected rates of in- and out-flow to all other occupations after adjustments for the marginals and diagonals of the mobility table. The notion of conditional stochastic equivalence is more general than that of clustering. Indeed, it can be shown that it subsumes clustering as a special case. Further, occupations that are conditonally stochastically equivalent satisfy the "homogeneity" criterion proposed by Goodman (1981) to determine the collapsibility of occupations into broader categories as well as the "internal homogeneity thesis" that was used by Breiger (1981) to identify social classes from worker flows. Partitions

recovered by community detection or clustering algorithms, on the other hand, do not satisfy these conditions.

While the formulation of DCSBMs is relatively simple, estimating the parameters of these models is computationally challenging: both the marginal likelihood of the observation model and the posterior distribution of the latent class memberships are generally intractable (Snijders and Nowicki 1997). This challenge is overcome by using a variational Expectation Maximization algorithm (VEM) that maximizes a tractable lower bound of the marginal log-likelihood to approximate the MLE of the model parameters (Daudin et al. 2008; Mariadassou et al. 2010). Simulation results show that the DCSBM fitted via the VEM algorithm is able to successfully recover (conditionally) stochastically equivalent positions from mobility tables under reasonable conditions. Further, it converges quickly and can be applied to mobility tables with more than 500 occupations and 10 classes within seconds.

For illustration, the model is fitted to two classical mobility tables: the mobility table analyzed in Breiger (1981) and the one analyzed in Goodman (1981). Both of these analyses show that the model is able to detect meaningful patterns of mobility, which are more general than clustering. Further, in both of these examples, the DCSBM gives good results, while most of the tested community detection algorithms fail to find any meaningful structure.

The paper unfolds as follows. First, two criteria that might be used to aggregate occupations based on mobility patterns, namely clustering and equivalence, are discussed and compared. Thereafter, a degree-corrected stochastic blockmodel is formulated together with an estimation algorithm that approximates the MLE of the model parameters. This is followed by a simulation study and two empirical applications. The paper concludes with a discussion of the results.

Equivalence and Clustering

Traditional approaches to analyzing mobility tables have relied heavily on log-linear models. These models assume that each cell count of the mobility table follows a Poisson distribution with a pre-specified rate parameter (Hout 1983; Agresti 2003). By constraining the parameter to interpretable patterns and testing the fit of the model to the data, researchers have been able to reach conclusions regarding the underlying pattern of occupational mobility in the population. While theoretically appealing, most parsimonious mobility structures—such as models of symmetry or quasi-independence—tend to fit observed data rather poorly, especially for larger mobility tables (Hauser 1978; Sobel et al. 1985). Furthermore, recent computational approaches have demonstrated that the structure of both inter- and intra-generational occupational mobility tends to be clustered within sets of occupations (Schmutte 2014; Toubøl and Larsen 2017; Cheng and Park 2020; Lin and Hung 2022), a pattern which is difficult to model with a log-linear models. Indeed, while "topological" models (Hauser 1978) are able to represent clustering, they require the researcher to manually specify the regions of homogeneous mobility flows without knowledge of the data, a task that becomes practically impossible once the mobility table becomes moderately large (Cheng and Park 2020).

Differently from the log-linear modeling tradition, in which researchers need to specify *a priori* a structure that is believed to capture the signal in the mobility table, recent computational approaches tend to adopt an inductive approach. Motivated by Weber's definition of "social class" as "the totality of those class situations within which individual and generational mobility is easy and typical" (Weber [1922]1978: 302), this stream of research takes the clustered nature of occupational mobility as a given and tries to identify the boundaries that separate occupations into classes defined by dense internal worker flows (Melamed 2015; Toubøl and Larsen 2017; Cheng and Park 2020). The criterion that these approaches optimize might be called the *Clustering Criterion* (CC, hereafter), as it reflects the extent to which the within-class worker flow is dense

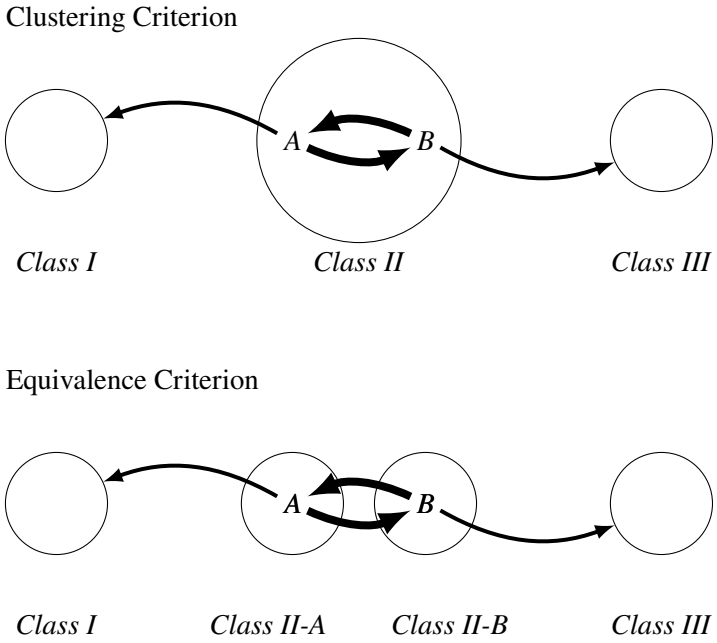
relative to that between classes. It is important to note that the CC does not aim at identifying *regions of cells* in the mobility table with homogeneous rates of mobility as the topological model does, but rather *sets of occupations* among which the worker flow is dense. Hence, the CC induces a simultaneous partition of the rows and columns of the mobility table, enabling the interpretation of the resulting aggregation scheme as a class structure, while the topological model, in general, does not share this property (Breiger 1981).

Equivalence vs Subgroups

Although the Clustering Criterion is, perhaps, the most intuitive way to aggregate occupations into classes, it is not the only one. Indeed, the earliest attempts that took the "aggregation question" (Breiger 1990: 8) of occupations into classes seriously focused on what we might call the *Equivalence Criterion* (EC, hereafter) (but see, Vanneman 1977), which aggregates occupations based on similarity in in- and out-flow patterns. Of particular relevance to this paper is the "internal homogeneity thesis" of Breiger (1981) and the "homogeneity" criterion discussed in Goodman (1981). Despite differences in the concrete model that they preferred, both Breiger and Goodman emphasized the *homogeneity* in in- and out-flow patterns between occupations as the major criterion to aggregate occupations into social classes. In other words, the property that needs to be shared among occupations in order to belong into the same class is not a high internal density of worker flow, but whether they send and receive workers at the same rate from the same occupations. Breiger (1981: 582) expressed this idea by stating that the class membership of occupations should "explain" the associations in the mobility table, while Goodman's homogeneity criterion reflects the same insight in that it requires origin and destination to be (quasi-) independent conditional on the classes to which they belong (Goodman 1981).

It is important to note that occupations that share equivalent in- and out-flow patterns

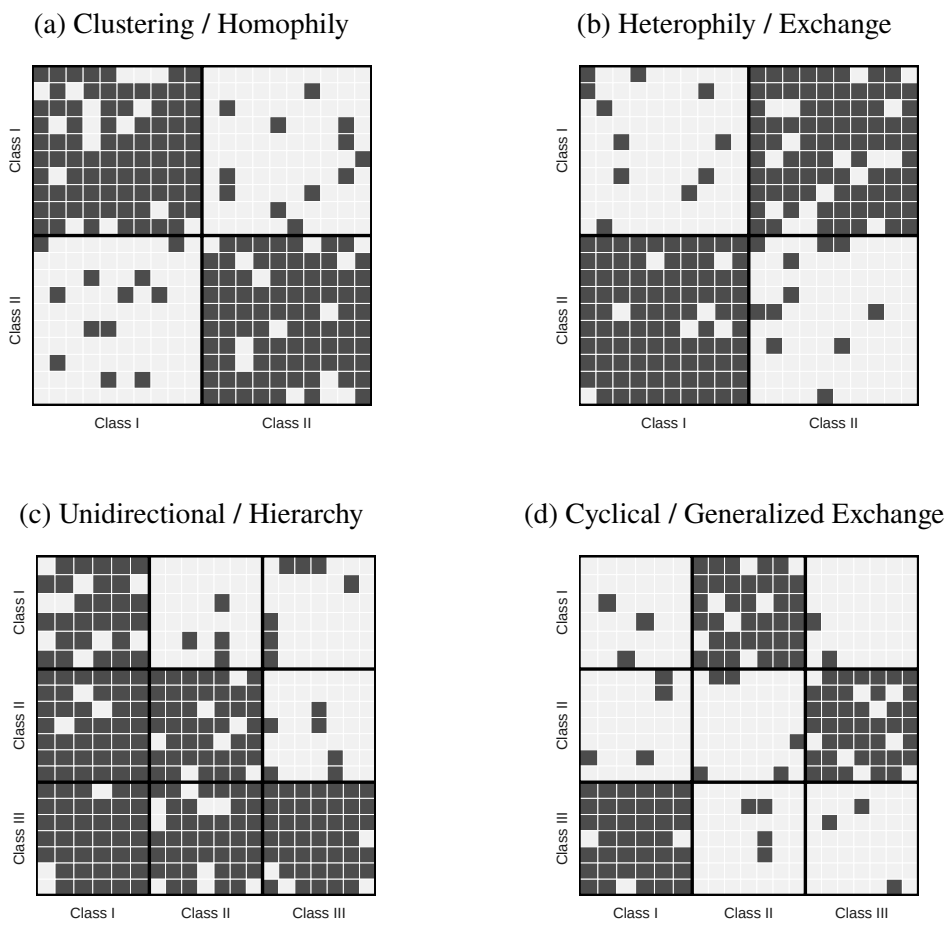
Figure 1: Illustration of Differences Between the Clustering Criterion and Equivalence Criterion in Aggregating Occupations into Classes



Note: Occupations are denoted by uppercase letters, while classes are represented by circles. Classes are assumed to contain multiple occupations but only *A* and *B* are highlighted for clarity. The arrows represent the direction and volume of the worker flow with thicker arrows reflecting higher volumes of workers.

do not need to have dense flows of workers among themselves, which shows that the EC is distinct from the CC. Figure 1 illustrates this by showing a scenario where the two criteria lead different groupings of occupations. The figure shows two occupations, *A* and *B*, between which large numbers of workers move. As the worker flow between the occupations is dense, the CC approach would allocate both into the same class, regardless of which other occupations *A* and *B* are connected to. The important point for the CC is that the worker flow between *A* and *B* exceeds that to other occupations or chance expectations based on the marginals of the mobility table. According to the EC, on the other hand, *A* and *B* are allocated to separate classes, because *A* (but not *B*) sends workers to occupations belonging to Class I, while *B* (but not *A*) sends them to occupations in Class III. Placing both *A* and *B* in the same class would violate the

Figure 2: Hypothetical Examples of Stochastically Equivalent Classes



Note: The rows in each table represent the origin, and the columns the destination, of hypothetical mobility flows. Dark cells represent origin-destination pairs of high mobility, while brighter cells represent those with low mobility. Black thick lines show the boundaries of stochastically equivalent classes.

EC, since the destination of the flow—i.e., occupations in Class I or III—depends on its origin—i.e., occupation *A* or *B*. Borrowing the terminology from the literature on social networks, we might say that the CC approach finds cohesive subgroups (Wasserman et al. 1994: Ch. 7) or community structures (Fortunato 2010), while the EC detects classes that are structurally equivalent (Lorrain and White 1971) or stochastically equivalent (Holland et al. 1983).

Although cohesive subgroups and structural equivalence have been, at times, pitted against each other in the literature on networks (e.g., Burt 1987; Erickson 1988), it

should be noted that equivalence is more general a criterion than cohesive subgroups in that the former is able to express a wider array of structures and subsumes clustering as a special case. To illustrate this point, Figure 2 shows four hypothetical mobility tables where the rows of the table represent the origin and the columns the destination of worker flows. Dark-colored cells represent pairs of occupations with high mobility, the light-colored cells those with little or no mobility, and the thick black lines the class boundaries. While all four examples in Figure 2 show meaningful mobility structures, the only pattern captured by the CC is (a), in which the within-class mobility flow exceeds that between classes. The EC approach, on the other hand, is able to capture all four structures. For instance, the two classes in Figure 2 (b) represent positions of approximately equivalent occupations, since all occupations in Class I tend to send and receive workers from Class II, and vice versa. In Figure 2 (c), all occupations in Class III send their workers to all other classes as well as their own; those in Class II send them to Class I and II, while workers in Class I circulate within their own class. Since all occupations belonging to the same class share the same mobility pattern, except for some random deviations, the three classes represent stochastically equivalent groups. Lastly, in Figure 2 (d), the worker flow between the three classes forms a cycle (I $\rightarrow$ II, II $\rightarrow$ III, and III $\rightarrow$ I), which is another case of stochastically equivalent classes which cannot be expressed as a clustering structure.

Of course, most observed mobility tables will not be as cleanly structured as the ideal-types presented here. Instead, these patterns should be regarded as examples of "sub-tables" or "local structures" that a model optimizing the EC is able to detect. It would not be surprising to find that some occupations occupy similar positions in the occupational system due to their exchange relation to another set of occupation, as in (b), while other groups of occupations are characterized by their dense within-class worker flow, as in (a). The EC encompasses all of these local configurations (and more) and, therefore, enables researchers to go beyond clustering in detecting structure from mobility tables.

Equivalence and Stochastic Blockmodels

In order to keep the presentation simple, three relations have been implicitly subsumed under the umbrella term *Equivalence Criterion*: stochastic equivalence (Holland et al. 1983), homogeneity (Goodman 1981), and internal homogeneity (Breiger 1981). All three relations share the intuition that similar or identical in- and out-flows should be the criterion to aggregate occupations into classes. However, they differ in their restrictiveness.

Stochastic equivalence requires that occupations belonging to the same class have the same probability distribution of in- and out-flows to all other occupations (Holland et al. 1983). This implies that all occupations of the same class must have the same expected in- and out-flow profile to all other occupations in the mobility table, irrespective of their sizes. Goodman's homogeneity criterion, on the other hand, treats a set X of occupations as *homogeneous* or *collapsible* if a model of quasi-independence fits the subtable created by considering only the rows and columns of the mobility table pertaining to occupations in X. For Poisson-distributed outcomes, this is equivalent to saying that occupations belonging to the same class share the same expected rate of in- and out-flows to all other occupations after adjustments for the marginals and diagonal cells of the mobility table. Hence, homogeneity can be understood as a *conditional* stochastic equivalence relation, where the mobility rates are conditioned on the total in- and out-flow as well as the number of stayers of each occupation. Lastly, Breiger's internal homogeneity thesis relaxes Goodman's homogeneity criterion further by adjusting the in- and out-flow probabilities of each occupation based on the marginal distribution of *each subtable* formed by crossing occupations belonging to one class with those belonging to another.[1] Hence, occupations that are stochastically equivalent

[1] To be precise, let y_{ij} be a discrete random variable that represents the count in the ith row and jth column of an $N \times N$ mobility table y, and let $\simeq$ denote the relation of stochastic equivalence. Two occupations, i and j, are said to be stochastically equivalent if the probability of any event defined on y remains unchanged when i and j are interchanged (Holland et al. 1983). So, $i \simeq j$ only if $p(y_{ih}) = p(y_{jh})$ and $p(y_{hi}) = p(y_{hj})$ for all $h \in O$, where O is the set of occupations. Notice that stochastic equivalence is indeed an equivalence relation and, thus, partitions the set O into M classes. Now, maintaining the

are homogeneous, and homogeneous occupations are internally homogeneous. Yet, the reverse direction of both of these implications are not necessarily true, and we might order the three relations as

$$\text{Stochastic Equivalence} \subseteq \text{Homogeneity} \subseteq \text{Internal Homogeneity}. \tag{1}$$

While homogeneity and internal homogeneity were formulated by mobility scholars, the notion of stochastic equivalence developed quite independently by networks scholars working on stochastic blockmodels (SBMs) (Holland et al. 1983; Wang and Wong 1987; Wasserman and Anderson 1987; Anderson et al. 1992). This might not be a coincidence: both of these developments can be understood as answering the call for a principled way to aggregate social entities based on relational information (White et al. 1976; Breiger 1990). And aggregating entities into classes based on the EC appears to be natural, since it minimizes the loss of relational information in the data (Marsden 1985). Indeed, occupations that are stochastically equivalent are, by definition, statistically indistinguishable/interchangeable with respect to their in- and out-flow probabilities. Hence, the class-by-class table created by aggregating them would loose no information regarding the relational signal contained in the original mobility table.

usual assumption of log-linear models that y_{ij} follows a Poisson distribution with parameter λ_{ij}, we note that λ_{ij} completely specifies the probability distribution in cell (i, j) of the mobility table and $i \simeq j$ only if $\lambda_{ih} = \lambda_{jh}$ and $\lambda_{hi} = \lambda_{hj}$ for all $h \in O$. Hence, the hypothesis that the mobility table y consists of M stochastically equivalent classes can be represented by the specification $\lambda_{ij} = \psi_{k[i]l[j]}$, where $\psi_{kl}, 1 \leq k, l \leq M$ is the expected frequency at which occupations in class k send workers to occupations in class l, and $k[i]$ and $l[j]$ denote, respectively, the classes to which occupations i and j belong. Notice that this model assumes that all occupations belonging to class k are expected to send the same number of workers to all occupations belonging to class l irrespective of their total in- and out-flow. The homogeneity criterion, on the other hand, requires only that the origin and destination are (quasi-)independent conditional on the class memberships, which is less restrictive. For a mobility table consisting of M homogeneous classes, $\lambda_{ij} = \alpha_i \beta_j \delta_{ij}^{\mathbb{I}(i=j)} \psi_{k[i]l[i]}$, where α_i, β_j, and $\delta_{ij}^{\mathbb{I}(i=j)}$ are, respectively, row-, column-, and diagonal-effects. Hence, stochastic equivalent classes are homogeneous, but not vice versa, and the former can be understood as a stronger version of latter with the constraints $\alpha_i = \beta_j = \delta_{ij} = 1$ for all $i, j \in O$. Lastly, Breiger's internal homogeneity thesis relaxes Goodman's requirement further by setting $\lambda_{ij} = \alpha_i^{(k,l)} \beta_j^{(k,l)} \delta_{ij}^{\mathbb{I}(i=j)} \psi_{k[i]l[i]}$, where $\alpha_i^{(k,l)}$ and $\beta_j^{(k,l)}$ are, respectively, row- and column-effects specific to the subtable formed by considering those occupations with origin in class k and destination in l (Breiger 1981: 589). Hence, Goodman's model of homogeneity can be understood as a model of internal homogeneity with the added constraints $\alpha_i^{(k,l)} = \alpha_i$ and $\beta_j^{(k,l)} = \beta_j$ for $1 \leq k, l \leq M$ and $i, j \in O$.

As similar these approaches are, so did mobility and network scholars share the same methodological difficulties. The most pressing challenge was the detection of class structures satisfying the EC *from* the observed data, instead of estimating parameters based on an *assumed* structure. Indeed, the introduction of methods to inductively recover stochastically equivalent classes from observed networks lead to a "rediscovery" of stochastic blockmodels in the statistical literature starting in the late 1990s (Snijders and Nowicki 1997; Nowicki and Snijders 2001; Daudin et al. 2008; Mariadassou et al. 2010; Karrer and Newman 2011; Zhang et al. 2015). Building on these developments, the next section formulates a degree-corrected SBM that is suitable for the analysis of mobility tables together with an estimation method—variational Expectation Maximization—that can be used to estimate the model parameters.

Stochastic Blockmodels for the Analysis of Mobility Tables

The degree-corrected stochastic blockmodel (DCSBM) used in this paper groups together occupations that are stochastically equivalent in their out- and in-flow pattern of workers after adjustments for occupations' total in- and out-flow as well as the number of stayers. The model can be motivated as a log-linear model with latent discrete variables that represent stochastically equivalent classes of occupations after these adjustments. In this section, the model is formulated together with a high-level introduction to the algorithm used to approximate the MLE of the model parameters. A more detailed discussion of the algorithm is delegated to the appendix.

The Model

Consider an $N \times N$ mobility table y, where N is the number of occupations, and where the (i, j)th cell of y corresponds to a random variable $y_{ij} \in \{0, 1, ...\}$ representing the

number of workers moving from origin occupation $i = 1, ..., N$ to destination occupation $j = 1, ..., N$. Following the log-linear modeling tradition, we assume that y_{ij} follows a Poisson distribution with parameter λ_{ij}. A convenient starting point to model λ_{ij} is to assume

$$\lambda_{ij} \equiv \alpha_i \beta_j \delta_{ij}^{\mathbb{I}(i=j)}, \tag{2}$$

where α_i and β_j are, respectively, the row- and column-effects that capture the total out- and in-flow of occupations, $\mathbb{I}(x)$ is an indicator function that is equal to 1 if x is true and 0 otherwise, and $\delta_{ij}^{\mathbb{I}(i=j)}$ are diagonal effects that capture the excess rate at which workers stay in their occupations. This model of "quasi-independence" (Goodman 1968) fits the diagonal cells of y perfectly, while assuming independence between origin and destination in the off-diagonal cells.

The DCSBM that is used throughout this paper assumes that any systematic deviations from the model of quasi-independence is explained by the class membership of the occupations. Assuming that there are M such classes, we might write

$$\lambda_{ij} \equiv \mu_{ij} \Psi_{z_i z_j}, \tag{3}$$

where $\mu_{ij} = \alpha_i \beta_j \delta_{ij}^{\mathbb{I}(i=j)}$ and $z_i \in \{1, 2, ..., M\}$ is a discrete latent variable that indicates the class to which occupation i belongs. Ψ_{kl} denotes the element in the kth row and lth column of an $M \times M$ *image matrix*, Ψ, that reflects the excess mobility rate from occupations belonging to class $k = 1, ..., M$ to those in class $l = 1, ..., M$ relative to the quasi-independence model.[2] Lastly, we assume that z_i is drawn independently from a categorical distribution with parameter π, where $\pi = [\pi_1, ..., \pi_M]^\top$ is a M-dimensional probability vector that satisfies $\pi_k > 0$ for all $1 \leq k \leq M$ and $\sum_{k=1}^{M} \pi_k = 1$, and where $x^\top$ denotes the transpose of x. This completes the model. The assumed data-generating

[2] All parameters are constrained to be positive. That is, $\alpha_i > 0, \beta_j > 0, \delta_{ij}^{\mathbb{I}(i=j)} > 0$, for all i, j; and $\Psi_{kl} > 0$ for all k, l. Further, to identify the model, the usual constraints $\prod_{i=1}^{N} \alpha_i = \prod_{j=1}^{N} \beta_j = \prod_{i=1}^{N} \delta_{ij}^{\mathbb{I}(i=j)} = 1$ are added. Notice that the elements of Ψ are allowed to vary freely because the model is formulated without a grand mean parameter.

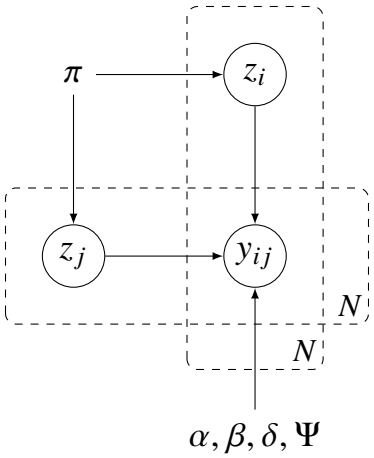
Figure 3: Data-Generating Process of Stochastic Blockmodel With Poisson Distributed Outcome and Adjustments for Row, Column, and Diagonal Effects

Degree-corrected Stochastic Blockmodel, DGP

Data: y
Parameters: $\theta = \{\alpha, \beta, \delta, \pi, \Psi\}$

for $i = 1, ..., N$ **do**
  $z_i \sim \text{Categorical}(\pi)$
end for

for $i = 1, ..., N$ **do**
  for $j = 1, ..., N$ **do**
    $\lambda_{ij} \leftarrow \alpha_i \beta_j \delta_{ij}^{\mathbb{I}(i=j)} \Psi_{z_i z_j}$
    $y_{ij} \sim \text{Poisson}(\lambda_{ij})$
  end for
end for



process (DGP) of the model is depicted in Figure 3 both as an algorithm and a directed acyclic graph.

Notice what the model postulates: after adjusting for the marginal in- and out-flow as well as the number of stayers of each occupation, the rest of the mobility table is completely characterized by the class membership of the occupations. In other words, it is assumed that the class structure "explains" the associations in the mobility table beyond quasi-independence, which is a strong version of the criterion proposed by Breiger (1981) to aggregate occupations into social classes. Reversely, once adjusted for the class memberships of occupations, the mobility pattern in y is quasi-independent, which is the homogeneity criterion of Goodman (1981). Hence, finding classes of occupations that are stochastically equivalent conditional on the marginal and diagonal effects is equivalent to finding sets of occupations that are homogeneous/collapsible according to Goodman's criterion. As homogeneity is a special case of internal homogeneity, it follows that these classes will be internally homogeneous as well.

The main difference between the current model and the approach adopted by Breiger and Goodman is how the class membership of the occupations, z, is treated. In the

approach of Breiger and Goodman, the class assignment is treated as a fixed and known feature of the model under the tested null hypothesis. Model parameters are estimated conditional on the assumed class partition, after which a goodness-of-fit test is performed to assess whether the assumed partition constitutes a good model of the data. Hence, inference regarding z is performed indirectly, following the logic of hypothesis testing. In the DCSBM, on the other hand, z is treated as a latent random variable. Model parameters are estimated by averaging over the uncertainty in z—i.e., by maximizing the marginal likelihood—after which the class assignment is directly predicted based on its posterior distribution. Hence, while the goal of both approaches is to find a partition of the occupations according to the EC, differences in the assumed nature of z lead to different inference procedures.

The DCSBM in (3) is also similar to the model used by Karrer and Newman (2011) with two differences: first, while Karrer and Newman (2011) developed their model for symmetric networks, the model in (3) differentiates between the origin and destination of worker flows. Second, Karrer and Newman did not "block out" the diagonal cells of the weighted adjacency matrix. This is equivalent to assuming that the mobility table can be described by a model of independence conditional of the block structure—i.e., that not only the between-occupation mobility rates but also the rate of staying in the same occupation is the same for occupations belonging to the same class after adjustments for the total in- and out-flow of occupations. Karrer and Newman (2011) justify this modeling choice by assuming that the network is sparse—i.e., that the probability of self-loops becomes negligible as the number of nodes in the network grows. Although a reasonable assumption for many networks, it is doubtful whether it will apply to mobility tables where stayers tend to occupy a large share of the total mobility flow regardless of the size of the table. By adding a separate set of parameters for the diagonal cells, the current model allows for within-class heterogeneity in the rate of stayers and aligns model closer with how mobility tables have been analyzed in the sociological literature on occupational mobility.

Estimation

Although the formulation of the DCSBM is relatively simple, estimating its parameters from observed data is challenging. Maximum likelihood estimation requires the maximization of the marginal likelihood:

$$p_\theta(y) = \sum_{z \in \mathcal{Z}} p_\theta(y, z) \tag{4}$$

with respect to the parameter vector $\theta = \{\alpha, \beta, \delta, \Psi, \pi\}$, where $p_\theta(y, z)$ is the complete-data likelihood and $\mathcal{Z}$ is the set of values that z can assume with positive probability. The main challenge is that the right-hand side of equation (4) is a sum of M^N terms, which becomes quickly intractable. While Markov Chain Monte Carlo (MCMC) methods have been proposed to estimate z, sampling approaches suffer from the notorious label switching problem (Stephens 2000; Jasra et al. 2005), which limits the interpretability of the results when the number of estimated classes become large and relabeling algorithms are less likely to be successful (Nowicki and Snijders 2001). In this paper, I instead use a variational expectation-maximization (VEM) algorithm to approximate the MLE of the parameters of the model (Daudin et al. 2008; Mariadassou et al. 2010).

The VEM algorithm proceeds by the same steps as the original EM algorithm (Dempster et al. 1977) with one major difference. In the usual EM algorithm, the E-step consists of calculating the posterior distribution of the unobserved variables, $p_\theta(z \mid y)$. Unfortunately, for DCSBMs, this distribution cannot be calculated in reasonable time for all but the smallest mobility tables (Snijders and Nowicki 1997). The VEM algorithm tries to overcome this problem by substituting the usual E-step with a *variational* E-step, where the intractable distribution $p_\theta(z \mid y)$ is approximated via a tractable *variational distribution*. The quality of the approximation depends on how close the variational distribution is to the target distribution, and finding a good candidate distribution is the art of this approach. In this paper, I follow Daudin et al. (2008) and Mariadassou et al.

(2010) and use a mean-field approach, which approximates the posterior $p_\theta(z \mid y)$ by a distribution that assumes independence between the class labels of occupations. Despite its simplicity, the mean-field approach has found many successful applications in the finite mixture modeling of network, text, and other types of data (Blei et al. 2017; Lee and Wilkinson 2019). For computational efficiency, the VEM algorithm is coded in `C++` using the `Eigen` library (Guennebaud et al. 2010). More details regarding the VEM algorithm and steps to choose initial values can be found in the appendix.

Simulation Study

A series of simulations were conducted to examine how the model behaves under different conditions. The model was tested for recovery of the class membership vector, z, as well as computation time.

Design

The parameters of the simulations were set as follows: the number of occupations were set to $N \in \{50, 100, 500\}$ and, for each value of N, the number of classes was varied between $M \in \{2, 3, 5\}$, except that for $N = 500$ a scenario with $M = 10$ classes was added as well.[3] For each value of N, the logged row- and column-effects were simulated independently from a standard Normal distribution and remained fixed throughout all simulation runs. Three types of image matrices were specified corresponding to the three ideal types in Figure 2 noting that the "Exchange" structure corresponds to the "Cyclical" structure for a network with two classes. The parameter $\gamma \in \{1, 2, 3\}$ was added to vary the strength of the block structure by setting the "dense" blocks in $\Phi = \log(\Psi)$ to γ and the "sparse" blocks to $-\gamma$, where $\log(\Psi)$ is the element-wise log-transform of Ψ. For example, for $M = 3$ classes, the three image matrices on which the model is tested are:

[3] For smaller values of N, it was difficult to generate data with the specified number of classes when the class proportions were skewed—i.e. when N was small and the $\nu = .5$.

$$\Phi_{\text{sym}}(\gamma) = \begin{bmatrix} \gamma & -\gamma & -\gamma \\ -\gamma & \gamma & -\gamma \\ -\gamma & -\gamma & \gamma \end{bmatrix}, \Phi_{\text{cycl}}(\gamma) = \begin{bmatrix} -\gamma & \gamma & -\gamma \\ -\gamma & -\gamma & \gamma \\ \gamma & -\gamma & -\gamma \end{bmatrix}, \Phi_{\text{hier}}(\gamma) = \begin{bmatrix} \gamma & -\gamma & -\gamma \\ \gamma & \gamma & -\gamma \\ \gamma & \gamma & \gamma \end{bmatrix}$$

The number of occupations in each class was varied across simulation runs by setting $\pi_k \propto \nu^k$ for $k = 1, ..., M$, where $\nu = \{0.5, 0.75, 1\}$. When $\nu = 1$, the elements of π will be all equal to $1/M$ leading to approximately equally sized classes; for smaller values of ν, the distribution π becomes more skewed. For each combination of the parameters, 25 datasets were simulated according to the algorithm in Figure 3 to which the model is fitted with 20 different initial values.[4] In total, this leads to $3^5 + 3^3 = 270$ different combinations of the parameters $\{N, M, \Psi, \gamma, \nu\}$ and $270 \times 25 = 6{,}750$ simulated datasets on which the model is tested. A summary of the full simulation design is shown in Figure A1 of the online supplement.[5]

Simulation Results

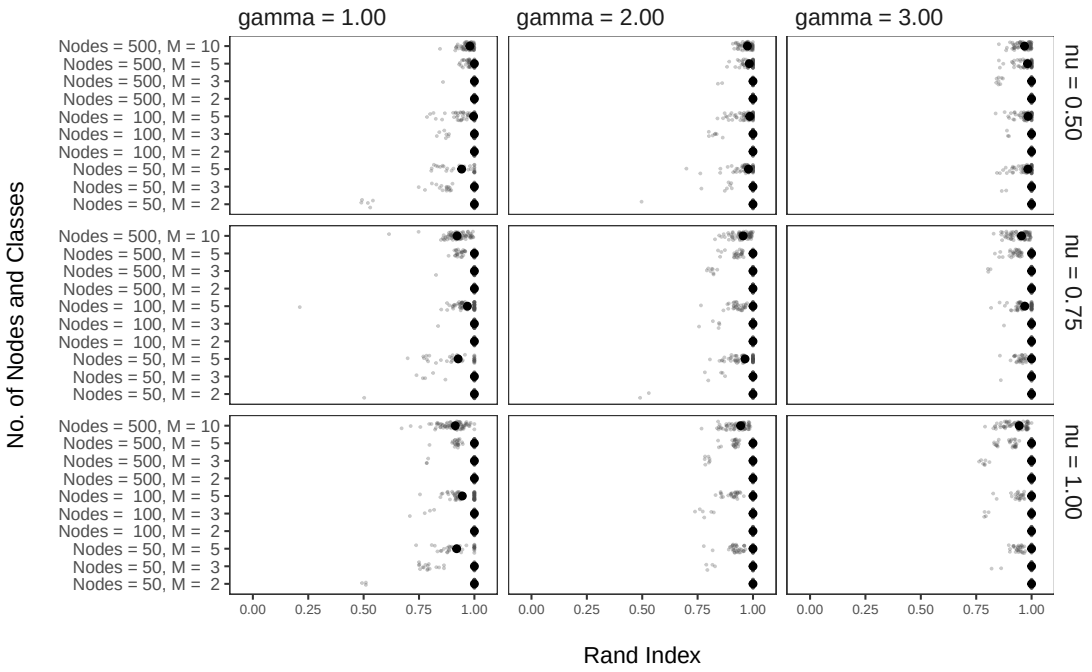
All models were run on the same machine equipped with an AMD EPYC 2.0GHz CPU using a single thread. For mobility tables with up to 100 occupations and 5 classes, fitting the model takes about a second or less. While the runtime increases slightly with the number of fitted classes when the number of occupations is relatively small, these differences are far less than those between mobility tables of small and larger sizes. Still, even with 500 occupations and 10 classes, the median run time was 20.8 seconds on a single thread, which shows that the model could be used for analyzing mobility tables of highly disaggregated occupational groups.[6] The full distribution of runtimes across

[4]The first six initial values are created by the spectral clustering algorithm described in the appendix on six different symmetrized mobility matrices: the simple symmetrization $y + y^\top$, bibliometric symmetrization, degree-discounted bibliometric symmetrization, and, for each variant, once using the graph Laplacian and once the symmetrically normalized graph Laplacian. If L is the graph Laplacian and D a diagonal matrix containing the degrees of the nodes, the symmetrically normalized graph Laplacian is defined as $L_{\text{sym}} = D^{-1/2}L^{-1/2}D^{-1/2}$. The rest of the initial values are created randomly.

[5]Notice that the diagonal parameters, $\delta_{ij}^{\mathbb{I}(i=j)}$ are not generated in the simulation because the diagonals are "blocked out" in the estimation process. The concrete parameters can be recovered noting that the diagonals of the mobility table are fitted perfectly.

[6]The 10th and 90th percentile of the runtime distribution for a model with 500 occupations and 10 classes were, respectively, 6.5 and 34.8 seconds. Leger (2016) reports that using the `R` package

Figure 4: Rand Index Between the "True" and Predicted Class Membership



Notes: Points are jittered vertically to show the distribution of the Rand index. Thick black dots represent the median of the distributions. Simulation runs that resulted in no variation in the class memberships are excluded from the figure. A corresponding plot of the normalized mutual information (see endnote 9 for the definition) can be found in Figure A3 of the online supplement.

different simulated scenarios are graphically represented in Figure A2 of the online supplement.

Figure 4 shows the distribution of the Rand index (Rand 1971) calculated between the "true" class membership vector, z, from which the data were simulated and the maximum a posteriori (MAP) prediction, $\hat{z}_{\text{MAP}}$, obtained after fitting the model. The thick black dot represents the median of the distribution, while the smaller dots represent the Rand Index for each individual run. The figure shows that the model is, in general,

`blockmodels` requires 85 seconds to fit a model *without* covariates to a graph with only 100 nodes and 10 classes; for a model that includes covariates, the reported runtime for a graph with 100 nodes and 10 classes is 2 hours 41 minutes and 16 seconds. These numbers are not directly comparable with those presented here, since the models are tested on different data, and all covariates in the model tested here are dummy variables. However, it is noticeable that the DCSBM could be fitted in less time than what the `blockmodels` package requires for a model without covariates on a smaller graph, since a closed-form solution exists for the M-step in the latter case while numerical methods have to be used for the DCSBM.

successful in recovering the "true" block structure. For almost all scenarios, the median Rand index is 1.00, indicating a perfect recovery of the class membership vector. Yet, the model performs better when the signal of the block structure is strong and when ratio of the number of occupations to that of classes is large. For example, the smallest median Rand Index across all simulation scenarios is found in the $\{N = 500, M = 10, \nu = 1.00, \gamma = 1.00\}$ and $\{N = 50, M = 5, \nu = 1.00, \gamma = 1.00\}$ scenario, both of which have low occupation-to-class ratios and the weakest block-signal ($\gamma = 1$). Still, even here, the median Rand Indices were both .915, showing a good match between the "true" and recovered class membership vector. Lastly, it should be noted that even in successful scenarios, we find that the algorithm sometimes stops at local maxima or degenerate solutions. This highlights the importance of starting the algorithm from multiple different initial values.[7]

Empirical Examples

To examine how the model performs on real data, two well-studied mobility tables were analyzed. The first example is the mobility table studied by Breiger (1981), which consists of intergenerational mobility patterns from father's occupation to son's first occupation collected in the Occupational Changes in a Generation (OCG-II) supplement to the March 1973 Current Population Survey. The dataset was originally analyzed by Featherman and Hauser (1978) and the occupations were pre-aggregated into 17 occupational groups by Breiger (1981) using the same categories as Blau and Duncan (1967). This example will be of importance, as it comes with a benchmark against which the partition recovered by the DCSBM can be compared to—namely Breiger's own partition. It is worth reiterating that the final model used by Breiger (1981) is not

[7] Out of the 6,750 runs, 188 resulted in a degenerate predictions of the class membership vector, where all occupations are allocated to the same class. Unsurprisingly, the degenerate results occurred when the table was small (93.1% of cases were found for the $N = 50$ scenario) and the "signal" of the block structure was weak (94.7% of the cases occurred when $\gamma = 1$, 5.3% occurred when $\gamma = 2$, and none when $\gamma = 3$). In Figure 4, these results are excluded.

exactly the same model as the DCSBM in equation (3) despite many similarities. Not only is the class membership vector, z, treated as known under the tested null hypothesis by Breiger but his final model (Hypothesis 7) fits separate row- and column-effects for each sub-table created by crossing one class with another, while the model in (3) fits only one set of row- and column-effects for each occupation.[8] Hence, the benchmark is not perfect. The second example analyzes the mobility data collected by Glass (2013[1954]), which was aggregated into an 8×8 mobility table by Miller (1960: 71), and analyzed by multiple mobility scholars including Duncan (1979) and Clogg (1981). This dataset was used by Goodman (1981) to demonstrate his approach to aggregating homogeneous occupations into classes. Hence, we might compare the partition inductively recovered by the DCSBM with that found by Goodman.

Comparison with Breiger (1981)

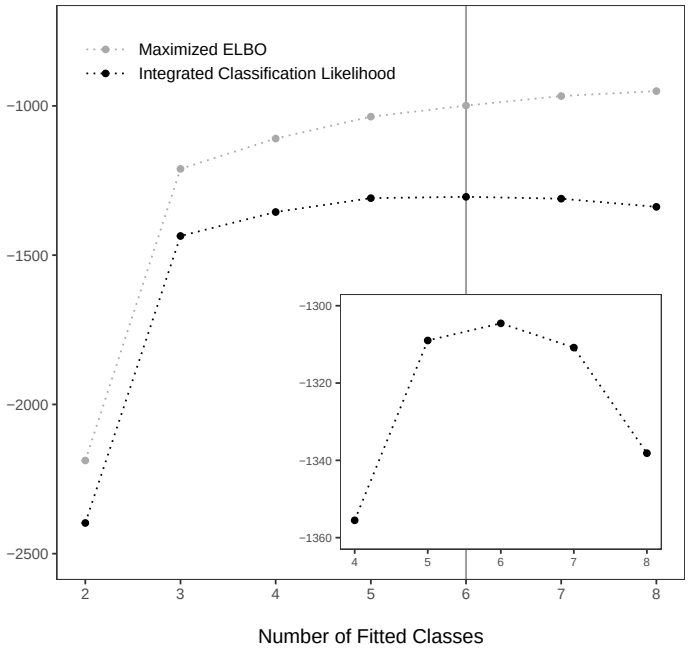
Models with 2 to 8 classes were fitted to the 17×17 mobility table analyzed by Breiger (1981). To select a final partition, the integrated classification likelihood (ICL) (Biernacki et al. 2000) was used, which is defined as

$$\text{ICL}(M, \theta) = \max_{\theta} \; \log p_\theta\Big(y, \hat{z}_{\text{MAP}}\Big) - \frac{1}{2}\Big[\log(N^2)P_1 + \log(N)P_2\Big] \tag{5}$$

where $\log p_\theta\Big(y, \hat{z}_{\text{MAP}}\Big)$ is complete-data log-likelihood evaluated at the MAP of z, $P_1 = 3(N-1) + M^2$, and $P_2 = M - 1$. The ICL criterion approximates the integrated complete-data log-likelihood via a Laplace approximation, similarly to how the Bayesian Information Criterion approximates the marginal log-likelihood (Raftery 1995). Differently to the BIC, however, the ICL lacks the "-2 multiplier" (which is added for historical reasons to put the statistic on the deviance scale) and, hence, models with

[8] In short, the DCSBM finds classes that are homogeneous (Goodman 1981), which is a stronger criterion than the internal homogeneity thesis (Breiger 1981) as discussed above. See footnote 1 for a formal comparison between these models. The implications of the differences have been discussed by Goodman (1981), Marsden (1985), and Hout (1983).

Figure 5: Integrated Classification Likelihood of DCSBMs fitted to the 17×17 Mobility Table in Breiger (1981)



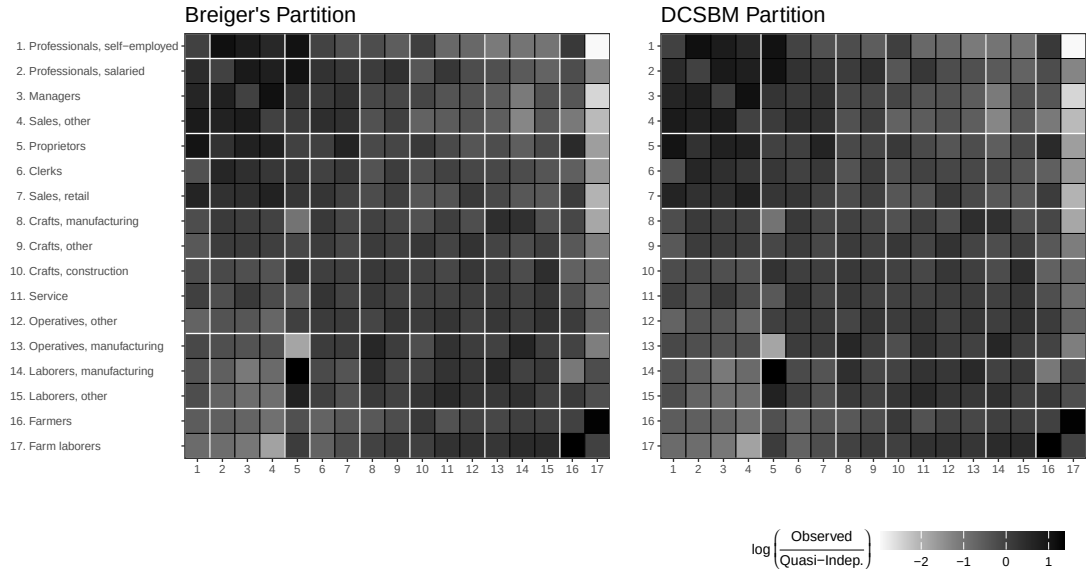
Notes: Models with higher ICL values are preferred. The ELBO is the lower bound of the marginal log-likelihood based on the variational distribution. The formal definition of the ELBO can be found in the appendix.

larger ICL values are preferred. Figure 5 shows the ICL values for models with two to eight classes with values from $M = 4$ to $M = 8$ magnified in the bottom-right of the plot. For each model, the VEM algorithm was started from 30 different initial values. According to the ICL criterion, the model with 6 classes fits the data best. Hence, in what follows, the 6-class model will be interpreted.

A comparison between the partition obtained by Breiger (1981) and the DCSBM is presented in Figure 6 and Table 1. Instead of showing the frequencies, the mobility table is represented as a heatmap in Figure 6. The colors represent the log-ratio $f^*_{ij} = \log(o_{ij}/e_{ij})$, where o_{ij} is the observed frequency in cell (i, j) of the table and e_{ij} is the expected frequency under a quasi-independence model with the diagonal entries of the cells blocked out.

Figure 6 shows that Breiger's partition and the DCSBM partition resemble each

Figure 6: Comparison of Block Structure Between DCSBM and Breiger (1981)



Note: Colors of the heatmap reflect values of $f^*_{ij} = \log(o_{ij}/e_{ij})$, where o_{ij} are the observed counts and e_{ij} are the expected counts of a quasi-independence model with the diagonal entries of the table blocked out. Darker colors indicate higher values of f^*_{ij}. The thick white lines show the block structure of the models, where the maximum a posteriori (MAP) estimate is used to allocated occupations into classes in the case of the DCSBM. The ordering of the occupations is the same in both plots. The Rand index and normalized mutual information between the partitions are, respectively, 0.93 and 0.87.

other quite closely. The Rand Index between the two partitions is 0.93, showing that over 90% of occupation-pairs are in agreement across the partitions in the sense that if two occupations are allocated in the same class in Breiger's partition, they are also in the same class of the DCSBM partition (and if they are in different classes in the former, they are also in different ones in the latter). While the Rand Index is easy to interpret, it doesn't take into account the marginal distribution of the two partitions. For this reason, the normalized mutual information (NMI) has been traditionally used to compare community structures in the network literature (e.g., Danon et al. 2005; Lancichinetti and Fortunato 2009; Yang et al. 2016). Calculated between the DCSBM and Breiger's partition, the NMI is 0.87 suggesting, again, a quite tight correspondence.[9] The only

[9]There are multiple ways in which the mutual information between two classification schemes can be normalized. Here, the arithmetic mean of the entropies is used for ease of comparison with earlier studies, i.e., $\text{NMI}(X,Y) = 2\text{MI}(X,Y)/[H(X) + H(Y)]$, where X and Y are (the probability distributions induced by) two classification schemes, $H(X)$ is the entropy of X, and $\text{MI}(X,Y)$ the mutual information

differences between the partitions are that "1. Professionals, self-employed" and "5. Proprietors" form their own class in Breiger's partition, while they are merged with adjacent occupations in the DCSBM-partition, and that "13. Operatives, manufacturing" are grouped together with the two laborer groups in Breiger's partition, while they are grouped with "12. Operatives, other," "11. Service," and "10. Crafts, construction" by the DCSBM. In short, it might be said that Breiger's partition is "almost" a refinement of the partition uncovered by the DCSBM.[10]

Perhaps most importantly, the results in Figure 6 show that the DCSBM detects partitions that make theoretical sense. Not only is the DCSBM partition similar to that of Breiger (1981), who had substantive knowledge of the mobility table and tried to find a partition satisfying a similar criterion as that optimized by the DCSBM, but the boundaries of the partition align with the order in which Blau and Duncan (1967) laid out the occupations, which, arguably, reflects their understanding of social distances between them. In short, the analysis shows that the DCSBM is able to detect meaningful patterns in mobility tables based on the flow of workers without supervision on part of the researcher.

Collapsibility

As the next step, the DCSBM partition is compared to the findings in Goodman (1981). The leading example in Goodman (1981) was the 8×8 mobility table shown in the right

between X and Y. $\mathrm{NMI}(X, Y)$ ranges from zero—when X and Y are independent or if $H(X)$ or $H(Y)$ is equal to zero—to one—which occurs when the classification schemes agree perfectly.

[10]It might be worth noting that Breiger's model provides a closer fit to the data than a log-linear model using the DCSBM partition. Breiger (1981) reports a G^2 statistic of 76.9 with residual degrees of freedom of 69 for his model (p. 596). As the log-likelihood of the saturated model, summing over the off-diagonals of the table, is $\ell_{\text{full}} = -714.205$, the log-likelihood of Breiger's model can be calculated as $-\frac{1}{2} \times G^2 + \ell_{\text{full}} = -752.655$ with $n(n-1) - df_{\text{resid}} = 272 - 69 = 203$ fitted parameters (see footnote 1 for an explicit formulation of Breiger's model). This is a higher log-likelihood than that resulting from a log-linear model based on the DCSBM-partition—i.e., when we use the specification in equation (3) with z substituted with $\hat{z}_{\text{MAP}}$—which is equal to −969.327 with 58 fitted parameters. Thus, based on the log-likelihood, Breiger's model offers a better fit to the data. This is not surprising, since Breiger's model fits more parameters to the table, and the log-likelihood is a nondecreasing function of model complexity. On the other hand, when the two models are compared using the BIC statistic, we obtain 2643.288 and 2263.791, respectively for Breiger's partition and the DCSBM-based partition. This might indicate that Breiger's model tends to overfit the data.

Table 1: Cross-tabulation of Partition Recovered from DCSBM and Partition Proposed by Breiger (1981)

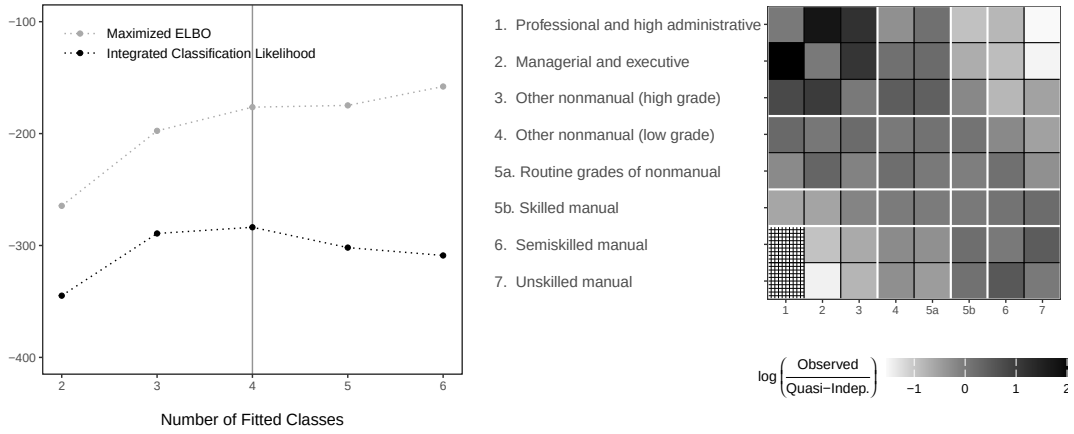
	DCSBM Partition					
Breiger (H7)	Class I	Class II	Class III	Class IV	Class V	Class VI
Class I	1					
Class II	3					
Class III		1				
Class IV		2				
Class V			2			
Class VI				3		
Class VII				1	2	
Class VIII						2

Note: Rows of the tables correspond to the eight classes of Breiger's (1981) final model and consist of Class I (1), Class II (2, 3, 4), Class III (5), Class IV (6, 7), Class V (8, 9), Class VI (10, 11, 12), Class VII (13, 14, 15), and Class VIII (16, 17), where the numbers in parentheses indicates the numbering of occupations as they appear in Figure 6. The columns of the table the classes recovered from the DCSBM, where the maximum a posteriori (MAP) estimates are used to allocated occupations into classes. The occupations belonging to each class are Class I (1,2,3,4), Class II (5, 6, 7), Class III (8, 9), Class IV (10, 11, 12, 13), Class V (14, 15), and Class VI (16, 17). Empty entries of the table are cells with zero counts. The Rand index and normalized mutual information between the partitions are, respectively, 0.93 and 0.87.

panel of Figure 7, where he found that occupations 4 and $5a$ should be collapsed, while $5a$ and $5b$ should belong into separate classes. As shown in the right panel of 7, the partition recovered by the DCSBM, where the number of classes are selected via the ICL, indeed groups occupations 4 and $5a$ together, while allocating $5b$ into different classes.

Yet, further analyses reveals that the classes recovered from the DCSBM might disagree with those based on Goodman's homogeneity test. In addition to $\{4, 5a\}$, the DCSBM partition collapses $\{1, 2, 3\}$ and $\{6, 7\}$ into classes as well, while only $\{6, 7\}$ passes Goodman's test as shown in Table 2. Interestingly, when Goodman's tests is applied to all possible pairs of occupations in the table, the results suggests that only $\{1, 2\}$ should be combined ($G^2 = 9.49, df = 11, p = .57$) instead of collapsing them together with $\{3\}$. The same tendency of collapsing more occupations than what Goodman's test would recommend is found when Breiger's table in Figure 6 is

Figure 7: Integrated Classification Likelihood and Recovered Partition from the DCSBM fitted to the 8×8 Mobility table in Goodman (1981)



(a) (b)

Note: Plot (a) shows the ICL statistic and the maximized ELBO for different number of fitted classes. The ELBO is the lower bound of the marginal log-likelihood based on the variational distribution. The formal definition of the ELBO can be found in the appendix. Plot (b) shows the mobility table analyzed by Goodman (1981) as a heatmap. Colors of the heatmap reflect values of $f^*_{ij} = \log(o_{ij}/e_{ij})$, where o_{ij} are the observed counts and e_{ij} are the expected counts based on a quasi-independence model with the diagonal entries of the table blocked out. Darker colors indicate higher values of f^*_{ij}, while the cell filled with a checked pattern shows indicates an undefined f^*_{ij} value. The thick white lines show the block structure recovered from an DCSBM with 4 fitted classes.

analyzed. Indeed, it turns out that none of the classes detected by the DCSBM pass Goodman's test, and only a single pair of occupations is found to be homogeneous—namely, "3. Managers" and "4. Sales, other." Hence, according to Goodman's criterion, the remaining occupations would be left as their own class. Notice that the DCSBM indeed collapsed the occupations that pass Goodman's test into the same class; however, instead of combining only $\{3, 4\}$, the DCSBM bundles $\{1, 2, 3, 4\}$ together.[11]

While it is difficult to reach general conclusions from two examples, these results suggests that the DCSBM detects partitions that are cruder than those resulting from the procedure recommended by Goodman (1981). Whether this behavior is desirable

[11]Full results of testing the collapsibility of all pairs of occupations in the mobility table presented in Figure 7b and Figure 6, respectively, can be found in Table A2 and Table A3 of the online supplement. Results of applying the homogeneity test to the classes detected by the DCSBM on Breiger's table are shown in Table A1.

Table 2: Goodness-of-fit Statistics of Quasi-independence Model Fitted to Classes Identified by the DCSBM Applied to the Mobility Table in Figure 7b

G^2	df	p	Collapsed Rows/Cols
48.41	21	0.00	1, 2, 3
7.87	11	0.73	4, 5
—	—	—	6
9.03	11	0.62	7, 8

Notes: G^2, df, and p are, respectively, the likelihood ratio chi-squared statistic, the residual degrees of freedom, and the associated p-value of the goodness-of-fit test when a quasi-independence model is fitted to only those rows and columns indicated in the fourth column of the table. Test results for the DCSBM partition of Breiger's(1981) table are shown in Table A1 of the online supplement.

will depend on the application. For example, a criterion that collapses only occupations 3 and 4 in Breiger's 17×17 mobility table might be considered too restrictive to be practically useful. Indeed, a log-linear model fitted to Breiger's table collapsing only those occupations that pass Goodman's test results in a higher BIC value than a model that collapses occupations according to the recovered DCSBM partition (see Table A4 of the online supplement). Perhaps practically more important, finding a well-fitting partition by testing the collapsibility of all possible subsets of occupations is computationally intractable except for very small mobility tables, since the number of models that need to be fitted increases exponentially in the number of occupations. Even for tables of dimension 17×17, such a procedure would require fitting $2^{17} - 1 = 131{,}071$ Poisson GLMs in the worst-case; for a 30×30 table, the number would jump to 1,073,741,823. Hence, similar to Hauser's topological model (Hauser 1978), Goodman's procedure is practically infeasible when searching *inductively* for a partition that summarizes mobility patterns in moderate to large tables. On the other hand, it is exactly for these tasks that recent computational methods, particularly community detection algorithms, have shown the most promise. Hence, in the next section, I compare the partitions recovered from these algorithms to the DCSBM partition.

Comparison with Community Detection Algorithms

Community detection algorithms and the DCSBM offer different approaches to finding structure in relational data. While a thorough comparison would go beyond the scope of this paper, a couple of major differences might be noted. First, community detection algorithms are computational algorithms that optimize a criterion function without specifying a data-generating process (DGP). As these approaches are not model-based, the criterion functions tend to be chosen by heuristics instead of being derived from first principles (but see, Rosvall and Bergstrom 2008; Rosvall et al. 2009). The DCSBM, on the other hand, is a fully probabilistic model based on distributions belonging to the exponential family. This leads to estimation procedures grounded in statistical theory—such as maximum likelihood or Bayesian approaches—with optimal properties (Bickel and Chen 2009; Celisse et al. 2012; Bickel et al. 2013).[12] Second, community detection algorithms are designed exclusively for the detection of partitions based on the clustering criterion. Hence, even in mobility tables that contain strongly structured worker flows, community detection algorithms will fail to find meaningful patterns unless these flows are confined within clusters. The DCSBM, on the other hand, groups occupations based on the equivalence criterion and, thus, is able to detect both clustering as well as more general structures. The generality of the DCSBM comes, however, at the cost of higher computational complexity. Indeed, most community detection algorithms are orders of magnitude faster than the algorithms used to fit DCSBMs. Hence, if the primary goal is to detect clusters in large datasets, community detection methods offer greater efficiency.

To compare community detection algorithms and the DCSBM on mobility data, I fit five widely used community detection algorithms to the two tables presented in

[12] It should be noted that the properties of the variational EM algorithm remain an active area of research. Bickel et al. (2013) provides conditions under which the VEM estimator for stochastic blockmodels as well as their degree-corrected variants using a fully factorized variational distribution is asymptotically equivalent to the maximum likelihood estimator. Yet, their results were limited for binary, symmetric, and sparse networks.

Table 3: Community Detection Algorithms Applied to the 17 × 17 Mobility Table Analyzed in Breiger (1981)

Occupation	Infomap	Walktrap	Edge Betweenness	Louvain	Leiden
1. Professionals, self-employed	1	1	1	1	1
2. Professionals, salaried	1	1	2	1	1
3. Managers	1	1	3	1	1
4. Sales, other	1	1	4	1	1
5. Proprietors	1	1	2	1	1
6. Clerks	1	1	5	1	1
7. Sales, retail	1	1	2	1	1
8. Crafts, manufacturing	1	1	5	2	1
9. Crafts, other	1	1	5	2	1
10. Crafts, construction	1	1	5	2	1
11. Service	1	1	5	2	1
12. Operatives, other	1	1	5	2	1
13. Operatives, manufacturing	1	1	5	2	1
14. Laborers, manufacturing	1	1	5	2	1
15. Laborers, other	1	1	5	2	1
16. Farmers	1	2	5	3	1
17. Farm laborers	1	3	5	3	1

Notes: Numbers are used to indicate clusters but have no substantive meaning. For the Edge Betweenness, Louvain, and Leiden algorithm, which maximize modularity, a symmetrized version of the mobility table was used in the analysis (see endnote 13 for details).

the previous section, where the tables are treated as weighted adjacency matrices: the Infomap (Rosvall and Bergstrom 2008), the Walktrap (Pons and Latapy 2006), the modularity maximization algorithm using edge-betweenness (Newman and Girvan 2004), the Louvain algorithm (Blondel et al. 2008), and the Leiden algorithm (Traag et al. 2019).[13] Results in Table 3 show that almost all tested community detection methods fail to find meaningful structures in the 17×17 table analyzed by Breiger. The Walktrap and Edge Betweenness algorithm create multiple single-occupation communities while lumping more than half of the occupations into one big cluster. Further, the Infomap and Leiden algorithm fail to find any communities at all. Only the Louvain algorithm finds an interpretable three-community solution: one cluster containing occupations 1

[13] It should be noted that the modularity criterion was designed for undirected graphs and has limitations when applied to directed (or directed and weighted) graphs (Kim et al. 2010; Fortunato 2010). For this reason, the mobility table was first symmetrized as $y_{\text{sym}} = (y + y^{\top})/2$ and then treated as the adjacency matrix of an undirected weighted graph when applying the three algorithms that optimize modularity: namely, edge betweenness, Louvain, and Leiden. For the Infomap and Walktrap algorithms, the original mobility table was analyzed directly. For all algorithms, the default parameters in the `R` package `igraph` (version 2.0.3) were used.

to 7, another cluster of 8 to 15, and the last cluster of "16. Farmers" and "17. Farm laborers". Although not necessarily designed for small networks, all of these algorithms have been successfully applied to graphs of comparable sizes. Given that the Louvain algorithm successfully finds three clusters, it might be speculated that the denseness of the table, instead of its small size, creates difficulties for other algorithms.

It is interesting to note that both Breiger's and the DCSBM partition are nested within the Louvain partition. Indeed, it appears that optimizing the EC "breaks up" clusters into more fine-grained classes based on their between-cluster connection patterns. For example, the first community detected by the Louvain algorithm—consisting of occupations 1 to 7—is broken into two classes by the DCSBM—occupations 1 to 4 and occupations 5 to 7. While the first of these classes has a high within-class density, the internal-density of the second class is close to what is expected under quasi-independence. Indeed, what distinguishes Class 2 is not its within-cluster density, but rather its higher-than-expected in- and out-flow from and to occupations in Class 1. Thus, while community detection lumps the two classes into one cluster based on their strong connections, the DCSBM differentiates between occupations within the same cluster that have high internal density and those that have similar external connections.[14] Similar conclusions are reached when the algorithms are applied to the mobility table analyzed by Goodman (1981) (results can be found in Table A5 of the online supplement).

Summary and Discussion

Recently introduced computational approaches to the analysis of mobility tables have mainly focused on detecting clusters of occupations with dense internal mobility flows. Yet, clustering is not the only way in which occupational mobility might be structured.

[14]The average of the logged mobility ratios within Class 1 of the DCSBM, not including the diagonal cells—i.e., $\frac{1}{N_{B_1}(N_{B_1}-1)} \sum_{\substack{i,j \in B_1 \\ i \neq j}} \log f^*_{ij}$, where B_1 is the set of occupations belonging to class 1 and $N_{B_1} = |B_1|$—is 0.788. The corresponding number for Class 2 is 0.252, and the average logged mobility ratios *between* Class 1 and 2—i.e., $\frac{1}{N_{B_1} N_{B_2}} \sum_{i \in B_1 \wedge j \in B_2} \log f^*_{ij}$—is 0.403.

As an alternative way to aggregate occupations, this paper focused on the equivalence criterion—i.e., grouping occupations with the same or similar in- and out-flows to all other occupations. It was discussed how "stochastic equivalence" (Holland et al. 1983), the "homogeneity" criterion (Goodman 1981), and the thesis of "internal homogeneity" (Breiger 1981) can be understood as special cases fo this criterion. Further, the paper shows how a degree-corrected stochastic blockmodel (DCSBM) with Poisson-distributed outcomes enables the inductive detection of classes that are stochastically equivalent conditional on the total in- and out-flow as well as the number of stayers of each occupation, thereby allowing researchers to identify patterns of mobility that go beyond clustering. Simulation results showed that the DCSBM fitted via a variational EM algorithm is able to successfully recover classes that are connected by patterns of clustering, cycles, and unidirectional flows. While the VEM algorithm performed generally well, the recovery of the "true" class memberships appears to depend on the (1) strength of the signal of the block structure vis-à-vis that of the row- and column-effects and (2) ratio of the number of occupations to the number of classes, where both a stronger signal and a higher ratio leads to a better recovery. The estimation algorithm is quite efficient: the model can be fitted to mobility tables with 500 occupations and 10 classes in less than 30 seconds, on average. Perhaps more importantly, the empirical examples showed that the DCSBM is able to identify meaningful structures in mobility tables in situations where most tested community detection algorithms failed.

While sociologists pioneered the idea of structural equivalence and blockmodels (Lorrain and White 1971; White et al. 1976), recent developments of their stochastic counterparts have yet to find applications in mobility research despite a surge in computational approaches. In hindsight this is surprising as well as unfortunate. It is surprising, because the idea of applying blockmodels to mobility patterns is quite old (e.g., Breiger 1981; Goodman 1981; Padgett 1990) and the DCSBM appears to be a natural extension of the traditional log-linear model, the main workhorse of mobility scholars. Yet, while other social scientists utilized various forms of stochastic blockmodels to find latent

structures in the labor markets (e.g., Nimczik 2017; Norris Keiller 2020; Fogel and Modenesi 2023), geographic mobility patterns (e.g., Carlen et al. 2022) and proximity data (e.g., Yu et al. 2018), sociologists have been slow in adopting recent advances of these models.[15] On the other hand, the late adoption of stochastic blockmodels is quite unfortunate, since aggregating the data according to the equivalence criterion retains more relational information than doing so by clustering (Marsden 1985). Indeed, two occupations belonging to the same "cluster" might have very different connection patterns outside its boundaries, while stochastically equivalent occupations have, by definition, the exact same probability to be connected to all other occupations both within and outside their class. Stochastic blockmodels and their degree-corrected variants enable the detection of such equivalent positions in the web of worker flows, either conditional on the row- and column-marginals as well as the diagonals of the mobility table or unconditionally.

This is not to say that clustering is inferior to equivalence as a criterion for aggregating mobility data nor that DCSBMs should be always preferred over community detection algorithms. Whether occupations should be grouped based on clustering or equivalence would depend on the concrete research question that is pursued, and, as mentioned above, community detection algorithms tend to be more efficient than the procedures used to estimate stochastic blockmodels. Still, many social phenomena require us to go beyond clustering to create an adequate representation their structures. After all, the social world is more complex than a juxtaposition of islands. Stochastic blockmodels offer sociologists a principled way to go beyond clustering structures and, hence, one step further toward capturing such complexities in mobility flows.

[15] I thank the anonymous reviewer who pointed out that Lin and Hung (2022) used stochastic blockmodels as a sensitivity check in their analysis of mobility tables (see footnote 11 on page 1565). To the best of my knowledge, this is the only application of the model to the analysis of mobility tables in the sociology literature.

Appendix

Variational EM Algorithm for Degree-corrected Stochastic Blockmodels

In order to motivate the VEM algorithm for DCSBMs, it is useful to start from the original EM algorithm (Dempster et al. 1977). The EM algorithm is often introduced as a procedure that maximizes the expected complete-data log-likelihood (e.g., Casella and Berger 2002). Yet, to connect it with its variational variant, it helps to conceptualize both the E- and M-step of the algorithm as procedures that maximize a lower bound of the marginal log-likelihood.

Following Bishop (2006), we start with a general decomposition of the marginal log-likelihood into a lower bound and a Kullback-Leibler divergence term. Let $\theta = \{\alpha, \beta, \delta, \Psi, \pi\}$ be the model parameters and $p_\theta(z \mid y)$ the posterior distribution of the class membership vector. Consider an arbitrary distribution $q(z)$ over $\mathcal{Z}$, the set of values that z can assume with positive probability. Dividing both sides of the equality $p_\theta(y, z) = p_\theta(z \mid y) p_\theta(y)$ by $q(z)$ and taking the logarithm of both sides, we obtain

$$\log \left[\frac{p_\theta(y, z)}{q(z)} \right] = \log \left[\frac{p_\theta(z \mid y)}{q(z)} \right] + \log p_\theta(y). \tag{6}$$

Next, multiplying both sides by $q(z)$ and summing over $z \in \mathcal{Z}$ gives, after rearranging,

$$\log p_\theta(y) = \sum_z q(z) \log \left[\frac{p_\theta(y, z)}{q(z)} \right] - \sum_z q(z) \log \left[\frac{p_\theta(z \mid y)}{q(z)} \right], \tag{7}$$

since $\sum_z q(z) = 1$. The first term on the right-hand side of (7) can be written as

$$\begin{aligned} \mathcal{L}_\theta(q) &= \sum_z q(z) \log p_\theta(y, z) - \sum_z q(z) \log q(z) \\ &= \mathrm{E}_{z \sim q} \Big[\log p_\theta(y, z) \Big] + H(q) \end{aligned} \tag{8}$$

i.e., the sum of the expectation of the complete-data log-likelihood with respect to the

distribution q and the entropy of q, $H(q) = -\sum_z q(z) \log q(z)$. The second term in (7) is the Kullback-Leibler divergence from q to the posterior distribution, $p_\theta(\cdot \mid y)$. Hence, we might express the marginal log-likelihood as

$$\log p_\theta(y) = \mathcal{L}_\theta(q) + \mathrm{KL}\Big(q \;\Big\|\; p_\theta(\cdot \mid y)\Big). \tag{9}$$

It can be shown that $\mathrm{KL}(P \,\|\, Q) \geq 0$ for any two discrete distributions P and Q with equality if and only if $P(z) = Q(z)$ for all $z \in \mathcal{Z}$ (Grünwald 2007). Hence, we see that

$$\log p_\theta(y) \geq \mathcal{L}_\theta(q), \tag{10}$$

or, in words, that $\mathcal{L}_\theta(q)$ is a lower bound of the marginal log-likelihood.

Both the E-step and the M-step of the EM algorithm can be understood as procedures that maximize the lower bound, $\mathcal{L}_\theta(z)$, by iterating between updating q and θ. The E-step keeps θ fixed and finds the distribution q that maximizes $\mathcal{L}_\theta(q)$. Since $\mathcal{L}_\theta(q)$ and the KL-divergence term in (9) must sum to $\log p_\theta(y)$, maximizing $\mathcal{L}_\theta(q)$ with respect to q is equivalent to minimizing the KL-divergence term. Hence, the optimal choice of q is setting $q = p_\theta(\cdot \mid y)$ for which $\mathrm{KL}(q \,\|p_\theta(\cdot \mid y)) = 0$. The M-step, then, fixes the distribution q, and maximizes the lower bound $\mathcal{L}_\theta(q)$ with respect to θ. Since q and, accordingly, $H(q)$ remains fixed in this step, this is equivalent to maximizing the posterior expectation of the complete-data log-likelihood. Under reasonable conditions, iterating between the E- and M-step to update q and θ is guaranteed to find a local maximum of the marginal log-likelihood (Wu 1983).

The main problem with applying the EM algorithm to DCSBMs is that the posterior distribution $p_\theta(\cdot \mid y)$ is generally intractable (Snijders and Nowicki 1997; Daudin et al. 2008). The variational EM algorithm tries to overcome this problem by approximating the optimal but intractable distribution $p_\theta(\cdot \mid y)$ with a tractable *variational* distribution. Let $\tilde{q}_\xi$ be a family of distributions, indexed by the parameter ξ, which is used for the

approximation. Substituting $\tilde{q}_\xi$ into (9), the marginal log-likelihood becomes

$$\log p_\theta(y) = \underbrace{\mathrm{E}_{z\sim\tilde{q}_\xi}[\log p(y,z)] + H(\tilde{q}_\xi)}_{\mathcal{L}_\theta(\tilde{q}_\xi)} + \mathrm{KL}\Big(\tilde{q}_\xi \;\Big\|\; p_\theta(\cdot \mid y)\Big), \tag{11}$$

where the lower bound based on the approximate distribution, $\mathcal{L}_\theta(\tilde{q}_\xi)$, is often referred to as the *variational lower bound* or *evidence lower bound* (ELBO).

The VEM algorithm proceeds in similar steps as the original EM algorithm. In the *variational* E-step, $\mathcal{L}_\theta(\tilde{q}_\xi)$ is maximized with respect to the variational parameters ξ, which is equivalent to choosing the distribution $\tilde{q}_\xi$ that minimizes the KL-term in equation (11) and, hence, offers the best approximation to the posterior distribution.[16] In the M-step, the variational distribution $\tilde{q}_{\hat{\xi}}$ is treated as fixed and $\mathcal{L}_\theta(\tilde{q}_\xi)$ is maximized with respect to θ. Notice that by choosing $\tilde{q}_\xi \neq p_\theta(\cdot \mid y)$, the KL-divergence term in (11) will not vanish, and the maximizer of $\mathcal{L}_\theta(\tilde{q}_\xi)$ will be an approximate value of $\hat{\theta}_{\mathrm{MLE}}$ in finite samples. How close $\hat{\theta}_{\mathrm{VEM}}$ is to $\hat{\theta}_{\mathrm{MLE}}$ will depend on the choice of the family $\tilde{q}_\xi$. Here, I follow Daudin et al. (2008) and Mariadassou et al. (2010) and use the fully factorized family

$$\tilde{q}_\xi(z) = \prod_{i=1}^{N} \tilde{q}_{\xi_i}(z_i) \tag{12}$$

where $\tilde{q}_{\xi_i}(z_i)$ denotes the categorical distribution with parameter ξ_i and where $\xi_i = [\xi_{i1}, \xi_{i2}, ..., \xi_{iM}]^\top$ is a M-dimensional probability vector satisfying the constraints

$$\begin{aligned} \xi_{ik} &\geq 0, && i = 1, ..., N \text{ and } k = 1, ..., M, \\ \sum_{k=1}^{M} \xi_{ik} &= 1, && i = 1, ..., N. \end{aligned} \tag{13}$$

The approach of approximating an intractable joint distributions by assuming independence between components is called a *mean-field approximation* and is often used in finite mixture models for network and similar data structures (Airoldi et al. 2008; Blei

[16] Of course, here "best" should be understood as the best possible approximation within the family of distributions that are considered.

et al. 2017; Lee and Wilkinson 2019).

For the fully factorize variational distribution in (12), the ELBO can be written as

$$\begin{aligned}\mathcal{L}_\theta(\tilde{q}_\xi) &= \mathrm{E}_{z\sim\tilde{q}_\xi}[\log p_\theta(y,z)] + H(\tilde{q}_\xi) \\ &= \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N} y_{ij}\log(\alpha_i\beta_j) + \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\sum_{k=1}^{M}\sum_{l=1}^{M}\xi_{ik}\xi_{jl}(y_{ij}\log\Psi_{kl} - \alpha_i\beta_j\Psi_{kl}) \\ &\quad + \sum_{i=1}^{N}\sum_{k=1}^{M}\xi_{ik}\log\pi_k - \sum_{i=1}^{N}\sum_{k=1}^{M}\xi_{ik}\log\xi_{ik} + C\end{aligned} \tag{14}$$

where C is a constant not depending on the parameters of the model and where the summation in the first two terms run over all off-diagonal cells of the mobility table, since the diagonals are "blocked out." In the variational E-step, $\mathcal{L}_\theta(\tilde{q}_\xi)$ is maximized with respect to ξ subject to the constraints in (13). This can be done via the method of Lagrange multipliers, leading to the consistency condition

$$\log\hat{\xi}_{ik} = \sum_{\substack{j=1\\ j\neq i}}^{N}\sum_{l=1}^{M}\xi_{jl}\log\gamma_{ijkl} + \log\pi_k + C_i \tag{15}$$

at the maximized lower bound, where

$$\log\gamma_{ijkl} = (y_{ij}\log\Psi_{kl} - \alpha_i\beta_j\Psi_{kl}) + (y_{ji}\log\Psi_{lk} - \alpha_j\beta_i\Psi_{lk})$$

and C_i is a constant not depending on ξ_{ik}. The variational E-step updates each $\hat{\xi}_i$ by cycling through equation (15) for each occupation $i = 1, 2, ..., N$, which can be considered as a coordinate ascent algorithm that increases $\mathcal{L}_\theta(\tilde{q}_\xi)$ until a local optimum is reached (Bishop 2006; Blei et al. 2017).

In the M-step, $\hat{\xi}$ is kept fixed and $\mathcal{L}_\theta(\tilde{q}_{\hat{\xi}})$ is maximized with respect to θ. The maximizer of π has a closed-form solution

$$\hat{\pi}_k = \frac{1}{N}\sum_{i=1}^{N}\hat{\xi}_{ik}. \tag{16}$$

For the rest of the parameters, $\eta = \{\alpha, \beta, \Psi\}$, we notice that the only part of the ELBO that depends on η is the expectation of the complete-data log-likelihood. Further,

$$\begin{aligned} \mathrm{E}_{z\sim\tilde{q}_\xi}[\log p_\theta(y,z)] &= \mathrm{E}_{z\sim\tilde{q}_\xi}[\log p_\theta(y \mid z) + \log p_\theta(z)] \\ &= \sum_{k=1}^{M}\sum_{l=1}^{M}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N} \xi_{ik}\xi_{jl} \log f(y_{ij};\lambda_{ijkl}) + K \end{aligned} \tag{17}$$

where $f(y_{ij};\lambda_{ijkl})$ is the PMF of the Poisson distribution with parameter $\lambda_{ijkl} = \alpha_i\beta_j\Psi_{kl}$ evaluated at y_{ij} and K is a constant not depending on η. Hence,

$$\begin{aligned} \operatorname*{argmax}_\eta \mathcal{L}_\theta(\tilde{q}_\xi) &= \operatorname*{argmax}_\eta \sum_{k=1}^{M}\sum_{l=1}^{M}\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N} \xi_{ik}\xi_{jl}\left\{y_{ij}\log(\alpha_i\beta_j\Psi_{kl}) - \alpha_i\beta_j\Psi_{kl} - \log(y_{ij}!)\right\} \\ &= \operatorname*{argmax}_\eta \left\{ \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N} y_{ij}\left[\log(\alpha_i\beta_j) + \xi_i^\top \log(\Psi)\xi_j\right] \right. \\ &\qquad \left. - \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N} \alpha_i\beta_j\left(\xi_i^\top\Psi\xi_j\right) \right\}, \end{aligned}$$

where $\log(\Psi)$ is the element-wise log-transform of the matrix Ψ. I use the the limited-memory BFGS algorithm (Liu and Nocedal 1989) implemented in the `lbfgs++` library to maximize this function, although other numerical methods could be used as well (Mariadassou et al. 2010).[17]

Initial Values

The choice of initial values for the VEM algorithm is important, since, as for other finite mixture models, the objective function can be highly multi-modal. As the objective of the model is to find stochastically equivalent classes after adjustments for node-degrees and loops, a degree-discounted bibliometric symmetrization (Satuluri and Parthasarathy

[17]The `lbfgs++` library is freely available from `https://github.com/yixuan/LBFGSpp/`. The gradient of the variational lower bound can be found in the online supplement.

2011) is used as the default method to create a symmetric version of the mobility table. For a square matrix X, the degree-discounted bibliometric symmetrization is given as

$$\tilde{X} = D_{\text{out}}^{-1/2} X D_{\text{in}}^{-1/2} X^\top D_{\text{out}}^{-1/2} + D_{\text{in}}^{-1/2} X^\top D_{\text{out}}^{-1/2} X D_{\text{in}}^{-1/2}, \tag{18}$$

where D_{in} and D_{out} are diagonal matrices containing the column- and row-sums of X—i.e., $D_{\text{in}} = \text{diag}(1^\top X)$, $D_{\text{out}} = \text{diag}(X1)$, where 1 is a vector of ones of compatible length.[18] After symmetrization, $M - 1$ eigenvectors are extracted from the (weighted) graph Laplacian $L = \text{diag}(1^\top \tilde{X}) - \tilde{X}$, starting with the eigenvector corresponding to the second smallest eigenvalue. These vectors are, thereafter, used as inputs to a k-means algorithm to obtain initial values for the membership vector z. The `armadillo` library (Sanderson and Curtin 2016) is used for fast calculations of the eigenvectors and clustering.

References

Agresti, Alan. 2003. *Categorical data analysis*. John Wiley & Sons.

Airoldi, Edoardo M, David M Blei, Stephen E Fienberg, and Eric P Xing. 2008. "Mixed Membership Stochastic Blockmodels." *Journal of Machine Learning Research* 9:1981–2014.

Anderson, Carolyn J, Stanley Wasserman, and Katherine Faust. 1992. "Building stochastic blockmodels." *Social networks* 14:137–161.

Bickel, Peter, David Choi, Xiangyu Chang, and Hai Zhang. 2013. "Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels." *The Annals of Statistics* 41:1922–1943.

Bickel, Peter J. and Aiyou Chen. 2009. "A nonparametric view of network models and Newman–Girvan and other modularities." *Proceedings of the National Academy of Sciences* 106:21068–21073.

[18] If X is a mobility table, then $x_{ij} \geq 0$ for all i, j. Hence, the entry in the ith row and jth column of the matrix $A = XX^\top$ is $a_{ij} = \sum_k x_{ik} x_{jk}$, which is large when occupations i and j send many workers to the same destination and small otherwise. Similarly, the entries of matrix $B = X^\top X$ are large for occupations that receive many workers from the same source. Hence, pairs of occupations that share similar out- and in-flow patterns from other occupations will have large corresponding entries in $\hat{X} = A + B = XX^\top + X^\top X$. $\hat{X}$ is called the bibliometric symmetrization of X. $\tilde{X}$ can be understood as a scaled version of $\hat{X}$, where the scaling is done by the occupations' total in- and out-flows.

Biernacki, Christophe, Gilles Celeux, and Gérard Govaert. 2000. "Assessing a mixture model for clustering with the integrated completed likelihood." *IEEE transactions on pattern analysis and machine intelligence* 22:719–725.

Bishop, Christopher M. 2006. *Pattern recognition and machine learning*. Springer.

Blau, Peter and Otis Dudley Duncan. 1967. "The American Occupational Structure." *New York: John Wiley & Sons* .

Blei, David M, Alp Kucukelbir, and Jon D McAuliffe. 2017. "Variational inference: A review for statisticians." *Journal of the American statistical Association* 112:859–877.

Block, Per, Christoph Stadtfeld, and Garry Robins. 2022. "A statistical model for the analysis of mobility tables as weighted networks with an application to faculty hiring networks." *Social Networks* 68:264–278.

Blondel, Vincent D, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. "Fast unfolding of communities in large networks." *Journal of statistical mechanics: theory and experiment* 2008:P10008.

Breiger, Ronald L. 1981. "The social class structure of occupational mobility." *American Journal of Sociology* 87:578–611.

Breiger, Ronald L. 1990. *Social mobility and social structure*. Cambridge University Press.

Burt, Ronald S. 1987. "Social contagion and innovation: Cohesion versus structural equivalence." *American journal of Sociology* 92:1287–1335.

Carlen, Jane, Jaume de Dios Pont, Cassidy Mentus, Shyr-Shea Chang, Stephanie Wang, and Mason A Porter. 2022. "Role detection in bicycle-sharing networks using multilayer stochastic block models." *Network Science* 10:46–81.

Casella, George and Roger Berger. 2002. *Statistical inference*. Duxbury.

Celisse, Alain, Jean-Jacques Daudin, and Laurent Pierre. 2012. "Consistency of maximum-likelihood and variational estimators in the stochastic block model." *Electronic Journal of Statistics* 6:1847–1899.

Cheng, Siwei and Barum Park. 2020. "Flows and boundaries: A network approach to studying occupational mobility in the labor market." *American Journal of Sociology* 126:577–631.

Clogg, Clifford C. 1981. "Latent structure models of mobility." *American Journal of Sociology* 86:836–868.

Danon, Leon, Albert Diaz-Guilera, Jordi Duch, and Alex Arenas. 2005. "Comparing community structure identification." *Journal of statistical mechanics: Theory and experiment* 2005:P09008.

Daudin, J-J, Franck Picard, and Stéphane Robin. 2008. "A mixture model for random graphs." *Statistics and computing* 18:173–183.

Dempster, Arthur P, Nan M Laird, and Donald B Rubin. 1977. "Maximum likelihood from incomplete data via the EM algorithm." *Journal of the Royal Statistical Society: Series B (Methodological)* 39:1–22.

Duncan, Otis Dudley. 1979. "How destination depends on origin in the occupational mobility table." *American Journal of Sociology* 84:793–803.

Erickson, Bonnie H. 1988. "The relational basis of attitudes." pp. 99–121, In *Social Structures: A Network Approach*, edited by Barry Wellman and S. D. Berkowitz. Cambridge University Press.

Featherman, David L and Robert Mason Hauser. 1978. *Opportunity and change*. Academic Press. New York NY, USA.

Fogel, Jamie and Bernardo Modenesi. 2023. "What is a Labor Market? Classifying Workers and Jobs Using Network Theory."

Fortunato, Santo. 2010. "Community detection in graphs." *Physics Reports* 486:75–174.

Glass, David Victor. 2013[1954]. *Social mobility in Britain*. Routledge.

Goodman, Leo A. 1968. "The analysis of cross-classified data: Independence, quasi-independence, and interactions in contingency tables with or without missing entries: Ra Fisher memorial lecture." *Journal of the American Statistical Association* 63:1091–1131.

Goodman, Leo A. 1981. "Criteria for determining whether certain categories in a cross-classification table should be combined, with special reference to occupational categories in an occupational mobility table." *American Journal of Sociology* 87:612–650.

Grünwald, Peter D. 2007. *The minimum description length principle*. MIT press.

Guennebaud, Gaël, Benoît Jacob, et al. 2010. "Eigen v3." http://eigen.tuxfamily.org.

Hauser, Robert M. 1978. "A structural model of the mobility table." *Social Forces* 56:919–953.

Holland, Paul W, Kathryn Blackmond Laskey, and Samuel Leinhardt. 1983. "Stochastic blockmodels: First steps." *Social networks* 5:109–137.

Hout, Michael. 1983. *Analysing Mobility Tables*. Sage Beverly Hills, CA.

Jasra, Ajay, Chris C Holmes, and David A Stephens. 2005. "Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling." *Statistical Science* 20:50–67.

Karrer, Brian and Mark EJ Newman. 2011. "Stochastic blockmodels and community structure in networks." *Physical review E* 83:016107.

Kim, Youngdo, Seung-Woo Son, and Hawoong Jeong. 2010. "Finding communities in directed networks." *Physical Review E* 81:016103.

Lancichinetti, Andrea and Santo Fortunato. 2009. "Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities." *Physical Review E* 80:016118.

Lee, Clement and Darren J Wilkinson. 2019. "A review of stochastic block models and extensions for graph clustering." *Applied Network Science* 4:1–50.

Leger, Jean-Benoist. 2016. "Blockmodels: A R-package for estimating in Latent Block Model and Stochastic Block Model, with various probability functions, with or without covariates." *arXiv preprint arXiv:1602.07587* .

Lin, Ken-Hou and Koit Hung. 2022. "The Network Structure of Occupations: Fragmentation, Differentiation, and Contagion." *American Journal of Sociology* 127:1551–1601.

Liu, Dong C and Jorge Nocedal. 1989. "On the limited memory BFGS method for large scale optimization." *Mathematical programming* 45:503–528.

Lorrain, François and Harrison C White. 1971. "Structural equivalence of individuals in social networks." *Journal of Mathematical Sociology* 1:67–98.

Mariadassou, Mahendra, Stéphane Robin, and Corinne Vacher. 2010. "Uncovering latent structure in valued graphs: a variational approach." *The Annals of Applied Statistics* 4:715–742.

Marsden, Peter V. 1985. "Latent structure models for relationally defined social classes." *American Journal of Sociology* 90:1002–1021.

Melamed, David. 2015. "Communities of classes: A network approach to social mobility." *Research in Social Stratification and Mobility* 41:56–65.

Miller, Seymour Michael. 1960. "Comparative social mobility." *Current Sociology* 9:1–80.

Newman, Mark EJ and Michelle Girvan. 2004. "Finding and evaluating community structure in networks." *Physical review E* 69:026113.

Nimczik, Jan Sebastian. 2017. "Job Mobility Networks and Endogenous Labor Markets." Number E05-V3 in Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Labor Contracts and Markets I, Kiel, Hamburg. ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft.

Norris Keiller, Agnes. 2020. "Detecting labour submarkets from worker-mobility networks: A preliminary study." IFS Working Paper W20/30, London.

Nowicki, Krzysztof and Tom A B Snijders. 2001. "Estimation and prediction for stochastic blockstructures." *Journal of the American statistical association* 96:1077–1087.

Padgett, John F. 1990. "Mobility as control: Congressmen through committees." pp. 27–58, In *Social mobility and social structure*, edited by Ronald L Breiger. Cambridge University Press Cambridge.

Pons, Pascal and Matthieu Latapy. 2006. "Computing communities in large networks using random walks." In *J. Graph Algorithms Appl*. Citeseer.

Raftery, Adrian E. 1995. "Bayesian model selection in social research." *Sociological methodology* pp. 111–163.

Rand, William M. 1971. "Objective criteria for the evaluation of clustering methods." *Journal of the American Statistical association* 66:846–850.

Rosvall, Martin, Daniel Axelsson, and Carl T Bergstrom. 2009. "The map equation." *The European Physical Journal Special Topics* 178:13–23.

Rosvall, Martin and Carl T Bergstrom. 2008. "Maps of random walks on complex networks reveal community structure." *Proceedings of the National Academy of Sciences* 105:1118–1123.

Sanderson, Conrad and Ryan Curtin. 2016. "Armadillo: a template-based C++ library for linear algebra." *Journal of Open Source Software* 1:26.

Satuluri, Venu and Srinivasan Parthasarathy. 2011. "Symmetrizations for clustering directed graphs." In *Proceedings of the 14th International Conference on Extending Database Technology*, pp. 343–354.

Schmutte, Ian M. 2014. "Free to move? A network analytic approach for learning the limits to job mobility." *Labour Economics* 29:49–61.

Snijders, Tom AB and Krzysztof Nowicki. 1997. "Estimation and prediction for stochastic blockmodels for graphs with latent block structure." *Journal of classification* 14:75–100.

Sobel, Michael E, Michael Hout, and Otis Dudley Duncan. 1985. "Exchange, structure, and symmetry in occupational mobility." *American Journal of Sociology* 91:359–372.

Stephens, Matthew. 2000. "Dealing with label switching in mixture models." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 62:795–809.

Toubøl, Jonas and Anton Grau Larsen. 2017. "Mapping the social class structure: From occupational mobility to social class categories using network analysis." *Sociology* 51:1257–1276.

Traag, Vincent A, Ludo Waltman, and Nees Jan Van Eck. 2019. "From Louvain to Leiden: guaranteeing well-connected communities." *Scientific reports* 9:5233.

Vanneman, Reeve. 1977. "The occupational composition of American classes: Results from cluster analysis." *American Journal of Sociology* 82:783–807.

Wang, Yuchung J and George Y Wong. 1987. "Stochastic blockmodels for directed graphs." *Journal of the American Statistical Association* 82:8–19.

Wasserman, Stanley and Carolyn Anderson. 1987. "Stochastic a posteriori blockmodels: Construction and assessment." *Social networks* 9:1–36.

Wasserman, Stanley, Katherine Faust, et al. 1994. *Social network analysis: Methods and applications*. Cambridge university press.

Weber, Max. [1922]1978. *Economy and society*, Volume 1. Univ of California Press.

White, Harrison C, Scott A Boorman, and Ronald L Breiger. 1976. "Social structure from multiple networks. I. Blockmodels of roles and positions." *American journal of sociology* 81:730–780.

Wu, CF Jeff. 1983. "On the convergence properties of the EM algorithm." *The Annals of statistics* 11:95–103.

Yang, Zhao, René Algesheimer, and Claudio J Tessone. 2016. "A comparative analysis of community detection algorithms on artificial networks." *Scientific reports* 6:30750.

Yu, Lisha, William H Woodall, and Kwok-Leung Tsui. 2018. "Detecting node propensity changes in the dynamic degree corrected stochastic block model." *Social Networks* 54:209–227.

Zhang, Xiao, Travis Martin, and Mark EJ Newman. 2015. "Identification of core-periphery structure in networks." *Physical Review E* 91:032803.

Online Supplement

Gradient of Variational Lower-Bound

The gradient of $\mathcal{L}_\theta(\tilde{q}_\xi)$ with respect to the (logged) parameters is

$$\frac{\partial \mathcal{L}}{\partial \log \alpha_i} = \left(\sum_{j \neq i}^{N} y_{ij}\right) - \alpha_i \xi_i^\top \Psi \tilde{\xi}_i, \qquad \frac{\partial \mathcal{L}}{\partial \log \beta_j} = \left(\sum_{i \neq j}^{N} y_{ij}\right) - \beta_j \tilde{\xi}_j^\top \Psi \xi_j,$$
$$\frac{\partial \mathcal{L}}{\partial \log \Psi_{kl}} = \sum_{i=1}^{N}\sum_{j \neq i}^{N} \xi_{ik}\xi_{jl} y_{ij} - \Psi_{kl} \sum_{i=1}^{N}\sum_{j \neq i}^{N} \xi_{ik}\xi_{jl}\mu_{ij} \tag{19}$$

where $\tilde{\xi}_i = \sum_{j \neq i} \beta_j \xi_j$ and $\tilde{\xi}_j = \sum_{i \neq j} \alpha_i \xi_i$.

Supplementary Figures and Tables

Figure A1: Design of Simulation Study

Symbol	Interpretation
N	Number of occupations
M	Number of classes
y_{ij}	Worker flow from occupation i to occupation j
α_i	row (out-flow) effect of occupation i
β_j	column (in-flow) effect of occupation j
Ψ	image matrix
π	class proportions
z_i	class membership of occupation i
ν	Parameter governing the skewness of class sizes in simulation (smaller values leading to more skewed class sizes)
γ	Parameter governing strength of block signal in simulation (larger values leading to stronger signal of the block-structure)
type	Type of the simulated between-class mobility pattern (see Figure 2)

Simulation Design

for $N \in \{50, 100, 500\}$ **do**

  for $i = 1, \ldots, N$ **do**

    $\log \alpha_i, \log \beta_i \sim \text{Normal}(0, 1)$

  end for

  if $N = 500$ **then** $\mathcal{M} = \{2, 3, 5, 10\}$ **else** $\mathcal{M} = \{2, 3, 5\}$

  for $M \in \mathcal{M}$, $\nu \in \{0.5, 0.75, 1\}$, type $\in \{\text{symm, cycl, hier}\}$, $\gamma \in \{1, 2, 3\}$ **do**

    for $k = 1, \ldots, M$ **do**

      $\tilde{\pi}_k \leftarrow \nu^k$

    end for

    $\pi \leftarrow \tilde{\pi} / \sum_k \tilde{\pi}_k$.

    $\Psi \leftarrow \Psi_{\text{type}}(\gamma)$

    for s = 1, ..., 25 **do**

      for $i = 1, \ldots, N$ **do**

        $z_i \sim \text{Categorical}(\pi)$

      end for

      for $i = 1, \ldots, N$ **do**

        for $j \neq i$ **do**

          $\lambda_{ij} \leftarrow \alpha_i \beta_j \Psi_{z_i z_j}$

          $y_{ij} \sim \text{Poisson}(\lambda_{ij})$

        end for

      end for

      for $r = 1, \ldots, 20$ **do**

        Generate initial values, $\theta^{(r)}_{\text{inits}}$

        $\hat{\theta}^{(r)} \leftarrow \text{VEM}(y^{(s)}, \theta^{(r)}_{\text{inits}})$

      end for

      $\hat{\theta} \leftarrow \text{argmax}_r \text{ ELBO}(\hat{\theta}^{(r)})$

      $\hat{z} \leftarrow \text{argmax}_x \, p_{\hat{\theta}}(x \mid y)$

      Compare $\hat{z}$ and z (via Rand Index or NMI)

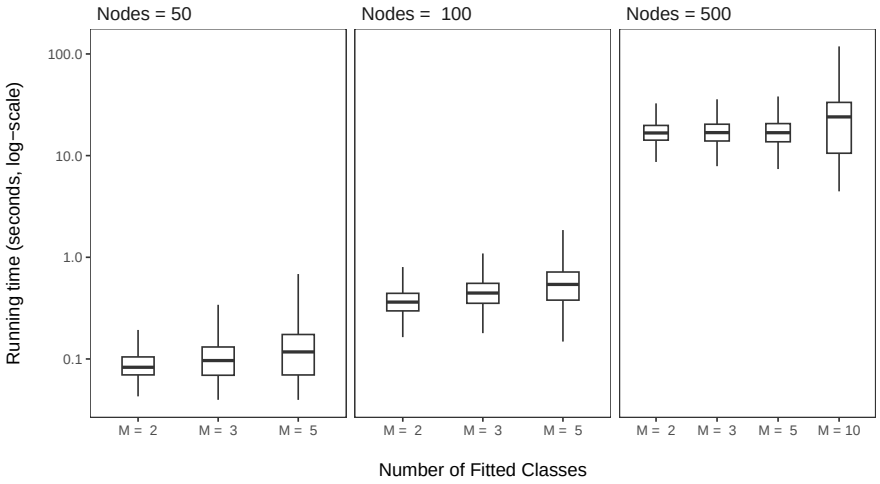
    end for

  end for

end for

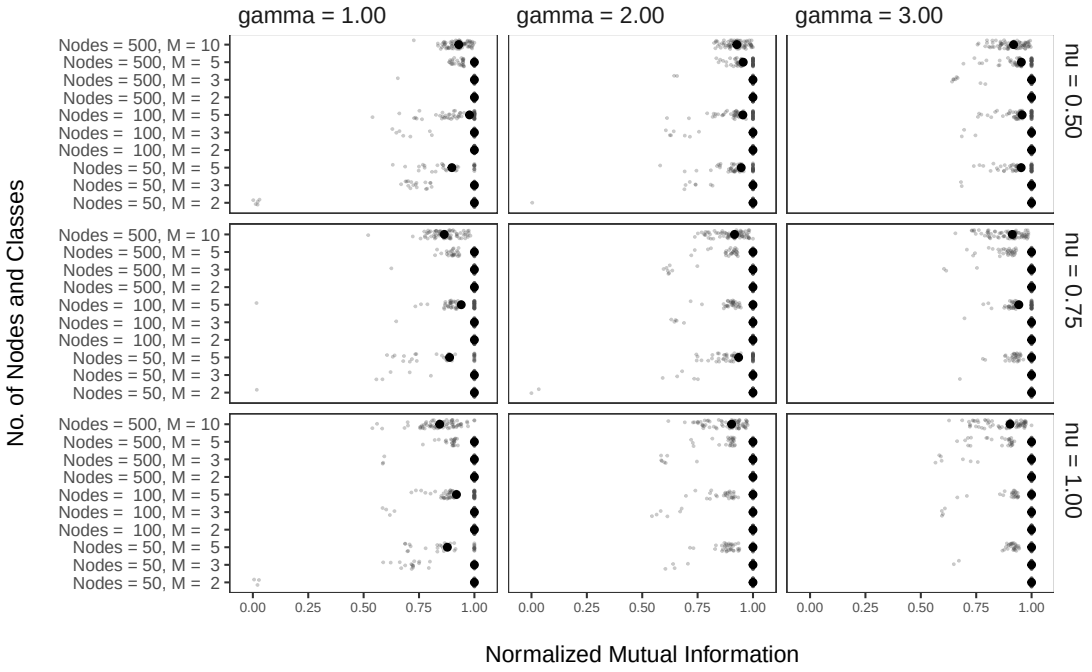
Note: Diagonal entries of y as well as their corresponding parameters, $\delta_{ij}^{\mathbb{I}(i=j)}$ are not simulated, since the model is fitted only to the off-diagonal elements of y.

Figure A2: Time to Fit the DCSBM to Simulated Data



Notes: The box show the first quartile, median, and third quartile of the distribution. The lines on top of each box extends to the largest value no further than 1.5 * IQR from the third quartile. The lines extending from the bottom of each box are defined analogously. All models were fitted using on a machine equipped with an AMD EPYC 2.0GHz CPU using a single thread.

Figure A3: Normalized Mutual Information of "True" and MAP Estimate of Class-membership Vector, z



Notes: Points are jittered horizontally to show the distribution of the normalized mutual information. Thick black dot represents the median of the distribution. Simulation runs that resulted in no variation in the block-memberships are excluded from the figure.

Table A1: Goodness-of-fit Statistics of Quasi-independence Model fitted to Classes Identified by DCSBM Fitted to the Mobility Table in Figure 6

G^2	df	p	Collapsed Rows/Cols
132.63	83	0.00	1, 2, 3, 4
112.70	57	0.00	5, 6, 7
105.32	29	0.00	8, 9
220.05	83	0.00	10, 11, 12, 13
84.62	29	0.00	14, 15
90.36	29	0.00	16, 17

Notes: G^2, df, and p are, respectively, the likelihood ratio chi-squared statistic, the residual degrees of freedom, and the associated p-value of the goodness-of-fit test when a quasi-independence model is fitted to only those rows and columns indicated in the fourth column of the table.

Table A2: Goodness-of-fit Statistics of Quasi-independence Models Fitted to All Pairs of Rows/Columns of Mobility Table in Figure 7b

G^2	df	p	Collapsed Rows/Cols
9.487	11	0.577	1, 2
40.987	11	0.000	1, 3
111.705	11	0.000	1, 4
89.552	11	0.000	1, 5
173.881	11	0.000	1, 6
205.871	11	0.000	1, 7
210.260	11	0.000	1, 8
27.906	11	0.003	2, 3
107.424	11	0.000	2, 4
83.893	11	0.000	2, 5
198.461	11	0.000	2, 6
249.532	11	0.000	2, 7
237.010	11	0.000	2, 8
65.155	11	0.000	3, 4
40.515	11	0.000	3, 5
208.683	11	0.000	3, 6
198.040	11	0.000	3, 7
230.224	11	0.000	3, 8
7.869	11	0.725	4, 5
75.589	11	0.000	4, 6
118.401	11	0.000	4, 7
112.112	11	0.000	4, 8
37.968	11	0.000	5, 6
78.000	11	0.000	5, 7
72.432	11	0.000	5, 8
39.475	11	0.000	6, 7
50.267	11	0.000	6, 8
9.028	11	0.619	7, 8

Notes: The numbers in the "Collapsed Rows/Cols" column indicate the row/column index of the occupational group. Hence, 5 corresponds to occupational group 5a in Goodman's table, 6 corresponds to 5b, 7 corresponds to 6, and 8 corresponds to 7.

Table A3: Goodness-of-fit Statistics of Quasi-independence Models Fitted to All Pairs of Rows/Columns of Mobility Table in Figure 6

G2	df	p	Collapsed Rows/Cols
42.810	29	0.047	1, 2
45.254	29	0.028	1, 3
44.205	29	0.035	1, 4
89.807	29	0.000	1, 5
85.372	29	0.000	1, 6
74.602	29	0.000	1, 7
184.881	29	0.000	1, 8
167.807	29	0.000	1, 9
228.712	29	0.000	1, 10
235.165	29	0.000	1, 11
278.213	29	0.000	1, 12
267.653	29	0.000	1, 13
300.900	29	0.000	1, 14
304.321	29	0.000	1, 15
424.766	29	0.000	1, 16
539.620	29	0.000	1, 17
52.901	29	0.004	2, 3
54.204	29	0.003	2, 4
63.220	29	0.000	2, 5
187.788	29	0.000	2, 6
102.297	29	0.000	2, 7
300.205	29	0.000	2, 8
265.603	29	0.000	2, 9
279.825	29	0.000	2, 10
332.191	29	0.000	2, 11
558.993	29	0.000	2, 12
662.259	29	0.000	2, 13
504.941	29	0.000	2, 14
553.460	29	0.000	2, 15
633.100	29	0.000	2, 16
2062.742	29	0.000	2, 17
17.622	29	0.952	3, 4
71.771	29	0.000	3, 5
131.338	29	0.000	3, 6
84.537	29	0.000	3, 7
333.334	29	0.000	3, 8
278.782	29	0.000	3, 9
363.909	29	0.000	3, 10
391.730	29	0.000	3, 11
547.887	29	0.000	3, 12
542.482	29	0.000	3, 13
461.211	29	0.000	3, 14
570.104	29	0.000	3, 15
1070.595	29	0.000	3, 16
1266.517	29	0.000	3, 17
69.117	29	0.000	4, 5
146.958	29	0.000	4, 6
97.312	29	0.000	4, 7
264.401	29	0.000	4, 8
248.793	29	0.000	4, 9
285.758	29	0.000	4, 10
311.831	29	0.000	4, 11
424.647	29	0.000	4, 12
458.017	29	0.000	4, 13
458.092	29	0.000	4, 14
489.164	29	0.000	4, 15
543.151	29	0.000	4, 16
1131.460	29	0.000	4, 17
71.233	29	0.000	5, 6
43.789	29	0.038	5, 7
150.404	29	0.000	5, 8
123.555	29	0.000	5, 9
167.220	29	0.000	5, 10
152.448	29	0.000	5, 11

219.799	29	0.000	5, 12
219.814	29	0.000	5, 13
217.523	29	0.000	5, 14
260.687	29	0.000	5, 15
752.159	29	0.000	5, 16
390.521	29	0.000	5, 17
51.377	29	0.006	6, 7
111.719	29	0.000	6, 8
97.974	29	0.000	6, 9
198.223	29	0.000	6, 10
167.151	29	0.000	6, 11
302.746	29	0.000	6, 12
323.047	29	0.000	6, 13
316.291	29	0.000	6, 14
356.467	29	0.000	6, 15
600.733	29	0.000	6, 16
1785.062	29	0.000	6, 17
95.976	29	0.000	7, 8
61.443	29	0.000	7, 9
119.056	29	0.000	7, 10
114.715	29	0.000	7, 11
199.194	29	0.000	7, 12
207.750	29	0.000	7, 13
240.366	29	0.000	7, 14
215.193	29	0.000	7, 15
358.948	29	0.000	7, 16
1066.349	29	0.000	7, 17
105.316	29	0.000	8, 9
147.583	29	0.000	8, 10
79.152	29	0.000	8, 11
131.664	29	0.000	8, 12
77.743	29	0.000	8, 13
130.461	29	0.000	8, 14
178.860	29	0.000	8, 15
788.194	29	0.000	8, 16
956.680	29	0.000	8, 17
61.758	29	0.000	9, 10
48.759	29	0.012	9, 11
88.994	29	0.000	9, 12
129.399	29	0.000	9, 13
152.266	29	0.000	9, 14
142.238	29	0.000	9, 15
683.437	29	0.000	9, 16
1023.436	29	0.000	9, 17
59.819	29	0.001	10, 11
45.239	29	0.028	10, 12
103.083	29	0.000	10, 13
107.906	29	0.000	10, 14
65.029	29	0.000	10, 15
484.023	29	0.000	10, 16
534.483	29	0.000	10, 17
55.406	29	0.002	11, 12
59.472	29	0.001	11, 13
82.302	29	0.000	11, 14
89.130	29	0.000	11, 15
537.793	29	0.000	11, 16
873.158	29	0.000	11, 17
109.931	29	0.000	12, 13
105.741	29	0.000	12, 14
48.062	29	0.014	12, 15
569.016	29	0.000	12, 16
1118.209	29	0.000	12, 17
48.029	29	0.015	13, 14
115.049	29	0.000	13, 15
736.901	29	0.000	13, 16
1315.631	29	0.000	13, 17
84.617	29	0.000	14, 15
202.540	29	0.000	14, 16
661.063	29	0.000	14, 17

297.277	29	0.000	15, 16
888.498	29	0.000	15, 17
90.358	29	0.000	16, 17

Notes: The numbers in the "Collapsed Rows/Cols" column indicate the row/column index of the occupational group.

Table A4: Bayesian Information Criteria for Different Models Fitted to Mobility Table in Figure 6

	No. of Classes	Log-likelihood	df	BIC
DCSBM	6	-1027.022	75	2479.027
Breiger	8	-1026.113	97	2601.870
Goodman	16	-780.711	260	3034.693
Quasi-Independence	1	-2783.245	50	5849.812

Notes: The Model named Goodman collapses "Managers" and "Sales, Other" into the same class, while leaving all other occupations as their own class. Notice that the statistics in the row named "Breiger" are different from those reported in Breiger (1981). This is because (1) only one set of row- and column-effects are fitted for all models and (2) the statistics reported in the table include the diagonals of the table as well as the parameters fitted to them. For the BIC statistic corresponding to Breiger's model as formulated in Breiger (1981) and fitted to only the off-diagonal cells of the table, see footnote 10 in the main text.

Table A5: Community Detection Algorithms Applied to Mobility Table Analyzed in Goodman (1981)

Occupation	Infomap	Walktrap	Edge Betweenness	Louvain	Leiden
1. Professional and high administrative	1	2	1	1	1
2. Managerial and executive	1	2	1	1	1
3. Other nonmanual (high grade)	1	1	1	1	1
4. Other nonmanual (low grade)	1	1	1	2	1
5a. Routine grades of nonmanual	1	1	1	2	1
5b. Skilled manual	1	1	1	2	1
6. Semiskilled manual	1	1	1	2	1
7. Unskilled manual	1	1	1	2	1

Notes: Numbers are used to indicate clusters but have no substantive meaning. For the Edge Betweenness, Louvain, and Leiden algorithm, which all maximize modularity, a symmetrized version of the mobility table was used in the analysis (see endnote 13).