Title

Dynamics of Population Flow Networks

Author

Huang, Peng

Publication Date

2024

UNIVERSITY OF CALIFORNIA,
IRVINE

Dynamics of Population Flow Networks

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Sociology

by

Peng Huang

Dissertation Committee:
Chancellor's Professor Carter T. Butts, Chair
Professor Veronica Berrocal
Professor Emerita Susan K. Brown
Professor David R. Schaefer

2024

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I write this dissertation on behalf of the many people who have supported me in making those scientific discoveries. First and foremost, I am deeply indebted to my advisor, Carter Butts, for generously sharing his knowledge, ways of thinking, and advice on every aspect of academic work. He is never absent when I am in need of guidance, and is always open to new ideas and different opinions. His advisorship has inspired me to constantly push myself out of the comfort zone to explore the unknowns, and think very hard to get ready for harsh but fair and thought-provoking debates. He shows me, with great passion in research and dedication to scientific rigor, what a scholar should be like.

I would also like to thank my dissertation committee members, David Schaefer, Susan Brown, and Veronica Berrocal, as well as my dissertation proposal committee member, Katherine Faust. They guide me through the mass literature in networks, demography, and statistics, nurturing and helping me refine my research ideas. I am deeply grateful to Wang Feng, who mentors me and stands for me in my navigation of the graduate school, the job market, and beyond. I thank Jacob Thomas for showing me the world of migration studies, and for the friendship and collaboration. I thank my spatial project collaborators, Loring Thomas, Yin Fan, Xiaoshuang Luo, Zack Almquist, and John Hipp. Working in this team has been fruitful, enjoyable, and educational.

The interdisciplinary academic training I received at UC Irvine across communities in sociology, statistics, and network analysis has greatly lifted my research. I am fortunate to learn, as a student or a teaching assistant, from Rachel Goldberg, Evan Schofer, Andrew Noymer, Nina Bandelj, Stanley Bailey, David Meyer, Zhaoxia Yu, Daniel Gillen, Bin Nan, Annie Qu, Volodymyr Minin, and Hal Stern. Numerous sociologists, demographers, and network scholars have generously provided me with advice, feedback, and encouragement, including Ashton Verdery, John Skvoretz, Ted Mouw, Filiz Garip, Jeffrey Johnson, Cassie McMillan, David Grusky, Andreas Wimmer, Liying Luo, Jennifer Van Hook, Victor Agadjanian, Tim Bruckner, Yu Xie, and Pavel Krivitsky. I am also grateful to Chunni Zhang, Zeqi Qiu, Qiong Wu, Biao Xiang, Xi Song, Yinxian Zhang, Zai Liang, and many other faculty and students who kindled my passion in social science research and supported my first few steps as a toddler researcher when I was an undergraduate student at Peking University.

My graduate study is filled with precious colleagueships and friendships with peers in the the sociology graduate program, the social network research group, the NCASD lab, the statistics brown-bag seminar committee, and others on and off campus. I am also indebted to my friends from college and childhood, who are thousands or tens of thousands miles apart but just a call away. I feel fortunate to share many joyful moments with you all.

I cannot express my gratitude enough to my family. I am grateful to my partner, who walked alongside me as we navigated graduate school together. Thank you for bearing with my anxiety, cheering for me, and bringing uncountable happiness to my life. I dedicate this dissertation to my parents, whose unwavering belief in me and unconditional love have been a constant source of strength.

Portions of Chapter 2 of this dissertation are a reprint of “Parameter Estimation Procedures for Exponential-family Random Graph Models on Count-valued Networks: A Comparative Simulation Study” in *Social Networks*, used with permission from Elsevier B.V.. The co-author listed in this publication, Carter Butts, directed and supervised research which forms the basis for the dissertation.

Portions of Chapter 3 of this dissertation are a reprint of “Rooted America: Immobility and Segregation of the Intercounty Migration Network” in *American Sociological Review*, used with permission from SAGE Publications. The co-author listed in this publication, Carter Butts, directed and supervised research which forms the basis for the dissertation.

Portions of Chapter 4 of this dissertation are a reprint of “California Exodus? A Network Model of Population Redistribution in the United States” in *Journal of Mathematical Sociology*, used with permission from Taylor & Francis. The co-author listed in this publication, Carter Butts, directed and supervised research which forms the basis for the dissertation.

This work was supported by the Geoffrey Tootell Dissertation-in-progress Award from the Mathematical Sociology Section of the American Sociological Association. It is also supported in part by the National Science Foundation under Grant SES-1826589 and the National Institutes of Health under Grant 1R01GM144964-01. The findings and conclusions expressed in this dissertation are those of the author and do not necessarily reflect the views of the aforementioned agencies.

VITA

Peng Huang

EMPLOYMENT

Assistant Professor of Sociology **2024-**
University of Georgia *Athens, GA, USA*

EDUCATION

Ph.D. in Sociology **2024**
University of California, Irvine *Irvine, CA, USA*

M.S. in Statistics **2023**
University of California, Irvine *Irvine, CA, USA*

M.A. in Sociology **2021**
University of California, Irvine *Irvine, CA, USA*

B.A. in Sociology **2018**
Peking University *Beijing, China*

B.A. in Economics **2018**
Peking University *Beijing, China*

FIELDS OF STUDY

Migration, Social Networks, Inequality, Population Health, Spatial Demography, Global and Transnational Sociology, Mathematical Sociology, Statistical and Computational Methods

REFEREED JOURNAL PUBLICATIONS

Huang, Peng, Carter T. Butts. 2024. "California Exodus? A Network Model of Population Redistribution in the United States." *Journal of Mathematical Sociology*, 48(3):311-339.

Huang, Peng, Carter T. Butts. 2024. "Parameter Estimation Procedures for Exponential-Family Random Graph Models on Count-Valued Networks: A Comparative Simulation Study." *Social Networks*, 76:51-67.

Thomas, Loring J., **Peng Huang**, Xiaoshuang Luo, John R. Hipp, Carter T. Butts. 2024. "Marginal-preserving Imputation of Three-way Array Data in Nested Structures, with Application to Small Areal Units." *Sociological Methodology*, 54(1):157-191.

Huang, Peng, Carter T. Butts. 2023. "Rooted America: Immobility and Segregation of the Intercounty Migration Network." *American Sociological Review*, 88(6):1031-1065.

Thomas, Loring J., **Peng Huang**, Fan Yin, Junlan Xu, Zack W. Almquist, John R. Hipp, Carter T. Butts. 2022. "Geographical Patterns of Social Cohesion Drive Disparities in COVID Infection Hazard." *Proceedings of the National Academy of Sciences (PNAS)*, 119(12):e2121675119.

Thomas, Loring J., **Peng Huang**, Fan Yin, Xiaoshuang Luo, Zack W. Almquist, John R. Hipp, Carter T. Butts. 2020. "Spatial Heterogeneity Can Lead to Substantial Local Variations in COVID-19 Timing and Severity." *Proceedings of the National Academy of Sciences (PNAS)*, 117(39):24180-24187.

GRANTS AND AWARDS

Conference Accelerator Grant **2023**
Institute for Humane Studies

Student Travel Award **2023**
International Network for Social Network Analysis

Geoffrey Tootell Dissertation-in-progress Award **2022**
Mathematical Sociology Section, American Sociological Association

Best Paper Award **2022**
Seventh U.S.-Japan Joint Conference on Mathematical Sociology and Rational Choice

ABSTRACT OF THE DISSERTATION

Dynamics of Population Flow Networks

By

Peng Huang

Doctor of Philosophy in Sociology

University of California, Irvine, 2024

Chancellor's Professor Carter T. Butts, Chair

Taking a relational and systemic approach, this dissertation offers theoretical, methodological, and empirical advancements in understanding the social forces that drive or inhibit human migration. We consider migration flows among geographical areas as a network system, analyzing its dynamics using the exponential-family random graph models (ERGMs) and simulation methods.

Chapter 2 grapples with the computational hurdle for modeling valued/weighted networks using ERGMs. We propose and implement an efficient parallelizable subsampled Maximum Pseudo-Likelihood Estimation (MPLE) scheme, which enables fast and accurate computation of ERGMs for big valued networks with high edge variance. The comparative simulation experiments further show whether and how the performance of existing computational approaches vary by the network size and the variance of edge values, providing guidelines for choosing and tuning those methods for different use cases.

Chapter 3 applies the implemented method to study intercounty migration in the United States (U.S.), whose migration rates have declined for decades and reached a historical low. We found a pattern of "segmented immobility," where fewer people migrate between counties with dissimilar political contexts, levels of urbanization, and racial compositions. We also propose a "knockout experiment" framework to quantify the impact of segmentation

on population immobility. The chapter reveals the social and political cleavages that underlie population immobility in the U.S., bridging the scholarly domains of (im)mobility, segregation, and polarization.

Motivated by debates about California's net migration loss ("California Exodus"), Chapter 4 examines the scale of and the mechanisms behind the migration-induced population redistribution among U.S. states. We combine ERGMs, knockout experiments, and a protocol for functional form visualization to understand the complex effects of political climates, housing costs, racial dynamics, and urbanization. The chapter offers an analytical framework for migration-induced population redistribution and demonstrates how generative statistical models can provide mechanistic insights beyond hypothesis-testing.

Chapter 1

Introduction

What drives people to migrate and what inhibits them from migrating? Where do migrants come from and where do they move towards? To tackle these fundamental questions about human migration, this dissertation conducts macrosociological research on the dynamics of population flows, using intercounty migration in the United States as the study case. We model the population flows between geographical locations as a network system, where the nodes/vertices represent places and edges represent the sizes of population migrating from one place to another.

Network analysis of migration flow systems offers three unique insights thanks to its three features: macroscopic, relational, and systemic. The majority of sociological research about migration happens at the micro-level, taking individuals and households as analytical units. Using qualitative and quantitative data collected mostly from migrant-sending regions, they identified how individual, household and sometimes neighborhood characteristics influence people's likelihood of migration, and the decision-making process involved (e.g. Chu 2010; Garip 2019; Liang et al. 2008; Lu et al. 2013; Massey et al. 1990; Massey and Espinosa 1997). Notwithstanding their rich discoveries, this well-developed approach, viewing through a wide-angle lens, is in essence case studies of migration flows, such as migration from Mexico to the U.S. or from rural to urban Thailand (Windzio 2018). They also tend to focus on cases that have extensive population flows, overlooking migration flows of smaller sizes and the mechanisms that hinder more migration from happening. As macrosociological research, this dissertation complements existing micro-level migration studies. We will analyze all of the 9.9 million possible intercounty migration routes in the U.S., avoiding selection biases towards major migration flow cases. The macro-level analysis also reveals how social contexts at larger geographical scales shape migration, such as the construct of state boundary, and the variation in political climates across counties. Furthermore, the dissertation will directly examine the key features of the whole migration system, including the size of internal migrant population in the U.S. and the scale of migration-induced population redistribution among U.S. states, as well as their social, economic, and political determinants.

The second feature of the following analysis is the relational perspective it takes. Extant macro-level research about migration, especially internal migration in the U.S., typically take geographical areas such as states and counties as research units (e.g., Greenwood, 1997; Partridge et al., 2012; Treyz et al., 1993). This approach is convenient and insightful about how local characteristics of a place can influence its migration rates. However, migration, which is population moving from one place to another, is by nature relational. This approach has to separate migration into two processes: immigration, and emigration, overlooking the possibly joint and interactive influence of origin and destination on migration. As Chapter 3 will show, the dissimilarity between counties play a critical role in influencing the scale of migration. This can only be discovered when taking a relational view, where migration is not a feature of one place (origin *or* destination), but a pair of places (origin *and* destination). Modeling the migration flow system as network operationalizes this perspective, where a migration flow is defined as an edge from its origin to its destination.

Lastly, as the migration systems theory has argued (de Haas, 2010; Mabogunje, 1970), population flows are not independent from each other, but form a system with internal dynamics, where the change in one migration flow can spillover to the other flows. This means that, statistically speaking, migration flows are autocorrelated, a feature that should be accounted for in order to draw valid inference. Moreover, the structure of the autocorrelation carries substantively meaningful information about the endogenous mechanisms of the migration flow system, such as whether and how migration flows are reciprocated (a larger flow from place A to place B promotes a larger flow in reverse from B to A). Network analysis offers a useful tool in understanding the systemic pattern of migration. Specifically, the exponential-family random graph models (ERGMs) is a parametric statistical tool that can explicitly specify the structure of interdependence of edges (migration flows) alongside other covariates that are exogenous to the network system, offering a path to reveal the internal dynamics of the migration flow system.

The macroscopic, relational, and systemic thinking has a long tradition in migration studies and can be traced back to the classic piece by Ravenstein (1885). Recent years have witnessed a rise in apply this thinking by using network methods, including ERGMs (e.g., Abramski et al., 2020; Akbari, 2021; DeWaard et al., 2020; Leal, 2021; McMillan, 2024; Windzio, 2018). The dissertation advances this inquiry in the following three areas:

First, computational methods for large-scale high-edge-variance valued network analysis using ERGMs. While network analysis can offer unique insights in studying migration flow systems, there exists one challenge. Most network methods are designed to model binary relationships, where an edge either exists or not at all. For migration systems and other flow systems, however, researchers are usually interested in more than whether there exist a flow at all, but the magnitude of the flow. This requires the model to be capable of handling networks of weighted/valued edges. Scholars have extended existing network methods for valued relationships (e.g., Opsahl et al., 2010), including ERGMS (Cranmer and Desmarais, 2011; Krivitsky, 2012; Krivitsky and Butts, 2017). Nevertheless, computation remains a hurdle for

ERGMs of valued networks (Valued ERGMs). As a result, network analyses of migration systems have to either be descriptive (Akbari, 2021), or perform data reduction (such as dichotomizing edges or putting them into quintiles) (Leal, 2021; McMillan, 2024; Windzio, 2018), losing information and running the risk of underestimating variability (Altman and Royston, 2006).

To address this issue, Chapter 2 proposes and implements an efficient parallelizable subsampled Maximum Pseudo-likelihood Estimation (MPLE) method for Valued ERGMs. We further test it against other existing computational methods via comparative simulation experiments. The experiments reveal how the performance of each parameter estimation method vary by the size of the network and the variance of edge value, providing guidance for choosing and tuning those methods for different tasks. Results show that MPLE offers accurate but much more efficient estimation of Valued ERGMs for high-edge-variance networks. This makes possible to model the quantitative features of large-scale network systems using Valued ERGMs, such as the inter-U.S.-county migration flow network with thousands of nodes and edge value ranging from zero to tens of thousands.

Second, macrosociological research about immobility. In recent years, scholars have realized the "mobility bias" in migration studies (Schewel, 2020), where researchers have extensively examined the social forces that drive migration, but largely neglected the counter forces that inhibit people from moving. Theoretical and empirical works on immobility are in a rise (e.g., Carling and Schewel, 2018; Liu and Peng, 2023; Mata-Codesal, 2015), yet less study immobility at the population and system level. This is especially relevant to the case of internal migration in the U.S., where migration rates have been declining for decades (Fischer, 2002; Molloy et al., 2011; Hyatt et al., 2018), but the reasons behind the population immobility is still inconclusive.

In response, Chapter 3 interrogates the lingering question about the social forces that limit migration, with an empirical focus on internal migration in the United States. We propose a systemic, network model of migration flows, combining demographic, economic, political, and geographic factors and network dependence structures that reflect the internal dynamics of migration systems. Using Valued ERGMs and MPLE developed in Chapter 2, we model the network of intercounty migration flows from 2011 to 2015. Our analysis reveals a pattern of "segmented immobility," where fewer people migrate between counties with dissimilar political contexts, levels of urbanization, and racial compositions. We further proposes a "knockout experiment" framework to quantify the impact of segmentation on population immobility. Results suggest that one would have observed approximately 4.6 million (27 percent) more intercounty migrants each year were the segmented immobility mechanisms inoperative. The chapter offers a systemic view of population immobility. It also reveals the social and political cleavages that underlie geographic immobility in the United States, showing the theoretical and empirical connections between immobility, segregation, and polarization from a relational perspective.

Lastly, investigation on the contribution of migration to population redistribution. Migration is a critical channel that fuels population redistribution across geographical space, with urbanization as a well-known and influential example (Lichter and Brown, 2011; Ravenstein, 1885). Nonetheless, migration does not always lead to population redistribution; population redistribution happens when migration is *asymmetric*, where the inflows and outflows of population are imbalanced. This means that the analytical framework useful for migration might not be effective when directly applied to study population redistribution. Researchers have developed measurements for and documented population redistribution via migration (Bell et al., 2002; Rees et al., 2017), but less is known about the driving forces behind and how to model them. This question is also empirically appealing, an example of which is the popular debates about "California Exodus" in public discourse (Bahnsen, 2021; Beam, 2021). Journalists and the public have been debating about the scale and causes of the net migration loss of California to other U.S. states, which has not yet received much research in demography or sociology.

Motivated by the empirical question about “California Exodus,” Chapter 4 develops analytical tools to model the mechanisms behind population redistribution. We introduce a protocol to visualize the complex effects on asymmetric flows in network systems, and advance the knockout experiment framework with positive and negative controls. Building on a modified model of the Valued ERGMs in Chapter 3, we identify the effects of political climate, urbanization level, racial composition, and housing costs on California’s net migration exchange with other U.S. states. Moreover, we also show that the severity of the California Exodus depends on how one measures it, and California is not the state with the most substantial population loss. The chapter demonstrates how generative statistical models can provide mechanistic insights beyond simple hypothesis-testing, and provides a framework to model migration-induced population redistribution and its underlying mechanisms.

Chapter 5 concludes the dissertation, summarizing the contributions, limitations, and directions for future work.

Chapter 2

Parameter Estimation Procedures for Exponential-family Random Graph Models on Count-Valued Networks: A Comparative Simulation Study

2.1 Introduction

Binary relations - relations in which edges can be approximated as simply “present” or “absent” - form the backbone of the social network field, with decades of theoretical, methodological, and empirical progress in understanding their structure and function. Valued relations, while by no means neglected, are less well-understood, and our tools for studying them less well-developed. Yet, the “strength of social ties” is at core of many scientific questions in a range of social settings (Granovetter, 1973; McMillan, 2022). Examples of valued relations include the frequency of interaction in interpersonal contact networks (Bernard et al., 1979), number of cosponsored bills shared among legislators (Cranmer and Desmarais, 2011; Fowler, 2006), communication volume within and among organizations (Drabek et al., 1981; Butts et al., 2007), encounters among non-human animals (Faust, 2011), and trade and migration flows among nations (Windzio, 2018; Ward et al., 2013). The need for rich information about social relations is particularly acute for networks involving interactions among aggregate entities such as nations, geographical areas, gangs, or formal organizations: because ties in such networks are themselves frequently aggregations of lower-level interactions, it is often the case that one's interest is not in the mere existence of trade, migration, homicide, communication, or other interactions, but their volume, frequency, or other quantitative features. In such settings, modeling edge states is of considerable substantive importance.

The earliest statistical modeling of valued relations was accomplished via network regression methods (Krackhardt, 1988); these provide only least-squares estimates of covariate effects, although autocorrelation-robust null hypothesis tests for such effects are well-known (Dekker et al., 2007), and some generalization via generalized linear models (GLMs) and related techniques is possible. Some forms of dependence can, further, be controlled semi-parametrically using latent structure models (e.g., Nowicki and Snijders, 2001; Hoff et al., 2002; Vu et al., 2013; Aicher et al., 2015), allowing estimation of covariate effects while accounting for unobserved mechanisms that can be written in terms of mixing on unobserved variables. Parametric models for valued graphs with general classes of dependence effects have been longer in coming, the current state of the art being exponential family random graph models (ERGMs) defined on sets of valued graphs (Block et al., 2022; Desmarais and Cranmer, 2012b; Krivitsky, 2012; Krivitsky and Butts, 2017); but see also Robins et al. (1999) for a pioneering example using categorical data and pseudo-likelihood estimation). Although ERGMs for valued graphs are not complete in the sense that they are for unvalued graphs (i.e., for most types of edge values, it is not always possible to write an arbitrary distribution on the order-N valued graphs in ERGM form), they are still highly general families, able to flexibly specify a wide range of effects. Since their introduction, they have been applied in a number of settings, ranging from networks of collaboration in government, and networks of migration flows, to networks of functional connectivity between brain regions (Huang and Butts, 2024a, 2023; Simpson et al., 2013; Ulibarri and Scott, 2017; Windzio, 2018).

Notwithstanding their broad applicability, parameter estimation for ERGMs in practice can be computationally demanding, a problem that is especially acute for valued networks. This issue has clearly had an impact on empirical network analyses in the published literature, forcing researchers to employ compromises or workarounds. As an example, Aksoy and Yıldırım (2024) noted in their paper that they could not obtain convergence for a single 81-node network using Valued ERGMs. For research that managed to obtain ERGM estimation of their valued networks, they had to either dichotomize the data and fall back to binary models (Leal, 2021), or coarsen the counts into quintiles (Windzio, 2018; Windzio et al., 2019); data transformation of this type greatly reduces computational difficulties, but in the meantime brings information loss and underestimation of variability (Altman and Royston, 2006). In short, even though methodological advances in valued network modeling have made it possible for researchers to capture quantitative features of relations beyond dichotomizational operations (Cranmer and Desmarais, 2011), the computational load remains a lingering hurdle to fully exploit the potential of these methods in scientific applications.

The major computational cost of ERGM estimation comes from the normalizing factor in its likelihood function, which is generally an intractable function involving the sum or integral of an exponentiated potential over the set of all possible network configurations. Although much is made over the fact that these sums have too many elements to explicitly evaluate (except in the case of extremely small unvalued graphs, e.g. Vega Yon et al., 2021), this is not the major obstacle to computation: rather, the difficulty rises from the extreme roughness (i.e., high variance) of the exponentiated potential over the support, which (in the absence of an explicit analytical solution) renders naive attempts at numerical approximation ineffective.

This problem can be amplified in the valued case, particularly where edge values vary greatly; valued edges can also pose challenges for some approximate estimation procedures that are successful in the case of unvalued ERGMs, as they must now explore a larger *per edge* state space. This high cost of estimation puts a priority on computationally efficient approximation methods. However, there has not been to date a systematic study of how well such methods perform, either in terms of improved computational efficiency or quality of estimation.

This chapter provides a look at this issue, evaluating estimation quality and computational cost for a number of alternative Valued ERGM estimation techniques. We focus on ERGMs for count-valued networks, i.e. relations whose edges take values on the unbounded non-negative integers, evaluating estimators via a simulation study based on intercounty migration-flow networks in two U.S. states. We vary the variance of edge values and the (node) size of the network, to simulate different data structures. The methods examined include the two currently implemented "standard" strategies - contrastive divergence (CD; Krivitsky, 2017) and Markov Chain Monte Carlo maximum likelihood estimation (MCMLE; Hunter et al., 2012) - as well as one approach not previously used in this setting, maximum pseudo-likelihood estimation (MPLE). MPLE is a workhorse approximation method in the binary ERGM case (Strauss and Ikeda, 1990), but requires special implementation measures for the count-data case, and to our knowledge has not previously been used for count-data ERGMs with general dependence. We also compare the performance of MPLE and CD as two seeding options for MCMLE. For all methods, we evaluate their computational speed, bias, variability, accuracy, calibration of estimated standard errors and confidence coverage.

The remainder of the chapter proceeds as follows. Section 2.2 briefly reviews ERGMs for valued networks, with applicable estimation strategies discussed in Section 2.3. Our simulation study design is described in Section 2.4, with results reported in Section 2.5. Section 2.6 discusses implications for method selection, and Section 2.7 concludes the chapter.

2.2 Count-valued ERGMs

An ERGM family for count data can be written as,

$$\Pr(Y = y|\theta, X) = \frac{h(y)\exp(\theta^T g(y, X))}{\sum_{y' \in \mathcal{Y}} h(y')\exp(\theta^T g(y', X))}, \tag{2.1}$$

where y is a realization of the network random variable Y on support $\mathcal{Y}$, the elements of which are graphs whose edges take values on the set $\{0, 1, \ldots\}$. (Here, we further assume that $\mathcal{Y}$ is a subset of the order-N count-valued graphs, though generalization is possible.) $g : \mathcal{Y}, X \mapsto \mathbb{R}^k$ is a vector of sufficient statistics, determined by exogenous covariates X and the graph state y, with corresponding parameter vector θ. Finally, $h : \mathcal{Y} \mapsto \mathbb{R}_{\geq 0}$ is the *reference measure*, which defines the limiting behavior of the model as $\theta \to 0$. Often tacitly taken to be constant for binary ERGMs, the reference measure is essential for Valued ERGMs, as it determines the marginal distribution of edge values under the reference (Krivitsky, 2012). Leaving $h(y) \propto 1$ leads to a marginal Boltzmann baseline distribution, while choosing

$$h(y) = \prod_{(i,j) \in \mathbb{Y}} (y_{ij}!)^{-1} \tag{2.2}$$

where $\mathbb{Y}$ is the set of edge variables, and y_{ij} is the value of the (i, j) edge, leads to a marginal Poisson baseline distribution of edge values. Other choices are also possible, some of which may have specific substantive interpretations (see e.g. Butts, 2019, 2020, for examples in the binary case).

As with binary ERGMs, we may specify the conditional probability that a given i, j edge variable will take a specified value. Again interpreting Y and y as random adjacency matrices, let Y_{ij}^c (respectively y_{ij}^c) refer to the set of all edge variables other than the ijth, and let the notation $z \cup Y_{ij}^c$ refer to the network formed by Y with the ijth edge variable set to value z.

Then we have

$$\Pr(Y_{ij} = y_{ij} | Y_{ij}^c = y_{ij}^c, \theta, X) = \frac{h(y_{ij} \cup y_{ij}^c) \exp(\theta^T g(y_{ij} \cup y_{ij}^c, X))}{\sum_{\ell=0}^{\infty} h(\ell \cup y_{ij}^c) \exp(\theta^T g(\ell \cup y_{ij}^c, X))} \tag{2.3}$$

$$= \left[\sum_{\ell=0}^{\infty} \frac{h(\ell \cup y_{ij}^c)}{h(y_{ij} \cup y_{ij}^c)} \exp\left[\theta^T \left(g(\ell \cup y_{ij}^c, X) - g(y_{ij} \cup y_{ij}^c, X)\right)\right] \right]^{-1}. \tag{2.4}$$

While the derivation is identical to the binary case (as can be appreciated by noting that Eq. 2.3 would reduce to the usual logistic form if ℓ were restricted to be ≤ 1), we note the computationally important difference that the conditional edge probability itself now has a non-trivial normalizing factor. In the general case, this has no analytical solution, and since it involves an infinite sum it cannot be explicitly evaluated otherwise. Although this does not impact e.g. the acceptance calculations for typical Markov Chain Monte Carlo (MCMC) algorithms (since the conditional odds of one graph versus another does not depend upon either normalizing factor), it does affect computation for the MPLE (which does depend on the conditional edge probability). Here, we formulate a finite sum approximation to Eq.2.3 for MPLE, as described below.

2.3 Estimation strategies for count-valued ERGMs

While many approaches to parameter estimation are possible, we focus here on approximations to the maximum likelihood estimator (MLE). Here, we briefly review the strategies employed, including special considerations for the count-data case. We note that some alternative schemes explored in the binary case (e.g., variational methods (Mele, 2017; Tan and Friel, 2020; Wainwright and Jordan, 2008)) may be adapted to the count data problem, but for purposes of this chapter we limit our study to approaches that have been established as broadly useful, and (with the exception of MPLE, which we extend) for which count-valued implementations currently exist.

2.3.1 Monte Carlo Maximum Likelihood Estimation

There are currently two widely used schemes for MCMC-based maximum likelihood estimation: stochastic approximation (Snijders, 2002; Wang et al., 2009), which is based on attempting to match the expected sufficient statistics to their observed values (exploiting the coincidence of methods-of-moments and MLE for exponential families); and the Geyer-Thompson method (Geyer and Thompson, 1992; Hunter et al., 2008b) (supplemented in current implementations by Hummel stepping (Hummel et al., 2012)), which uses an importance sampling scheme to directly optimize the log-likelihood surface. We here employ the former in its `statnet` implementation (Hunter et al., 2008b; Krivitsky et al., 2012).

MCMLE methods are the current gold-standard techniques for ERGM maximum likelihood estimation, with good theoretical properties (Snijders, 2002; Handcock, 2003) and strong performance in simulation studies for binary networks (van Duijn et al., 2009). An important bottleneck impacting the use of MCMLE, however, is the ability to produce relatively high-quality draws from the specified ERGM distribution (without which, the algorithms will not converge correctly). While it is known that conventional MCMC algorithms can in principle mix arbitrarily slowly (Snijders, 2002; Bhamidi et al., 2008), in actual practice this problem has been observed primarily in badly specified models that are degenerate or near-degenerate, and hence of limited relevance in typical social network applications (Hunter et al., 2012). That said, estimation time can still become long on very large networks, particularly for models with strong edgewise dependence.

This cost issue becomes more acute for Valued ERGMs, especially where edge values are highly variable. Intuitively, good MCMC mixing requires the Markov chain to explore the space of high-probability graphs, whose size increases substantially when edge values vary over a large range. For instance, for a simple random walk MCMC algorithm that proposes perturbing edges at random,[1] $\mathcal{O}(N^2)$ toggles may be needed to ensure that every edge variable in an unvalued graph has a high probability of having the "opportunity" to change state. If edges typically vary over some interval of order R, then a similar random walk scheme that increments or decrements edge values will need at least $\mathcal{O}(RN^2)$ for each edge to have the "opportunity" to cover its range of values. For networks with large counts (e.g., migrant-flow networks), one can easily obtain $R \gg N$, in which case simulation costs can rapidly become prohibitive. Although this problem can be alleviated by coarsening edge values to a much smaller range (as done e.g. by Windzio, 2018; Windzio et al., 2019), this both loses information and distorts the resulting model (since e.g., coarsening artificially reduces the entropy of the graph distribution). In principle, improved MCMC algorithms offer a better way to address this problem in the long-term, but current implementations do not seem to scale well for high-variance count models (e.g. Aksoy and Yıldırım, 2024). As we show below, MPLE can often deliver comparable estimation quality to MCMLE for high-R Valued ERGMs, where the latter suffers substantial increases in computational cost.

2.3.2 Contrastive Divergence

One alternative to either numerical approximation of expected statistics or of log likelihood ratios is to use a local approximation to the gradient of the likelihood in regions of the support "near" the observed data. This is the essential idea behind contrastive divergence (CD) (Hinton, 2002), a method originally introduced in the machine learning literature for scalable inference that is particularly well-suited to ERGMs and other exponential families (Asuncion et al., 2010; Krivitsky, 2017). CD can be employed for both valued and unvalued graphs, and greatly reduces computational time by using using only very short MCMC chains

---

[1] Practical implementations often use slightly different proposals, but the basic intuition carries.

starting at the observed data, depending on neither sample convergence nor burn-in. It is, however, an approximate technique that optimizes a function closely related to the pseudo-likelihood (Asuncion et al., 2010), and thus shares some of the drawbacks of the MPLE. These properties make CD a reasonable seeding method that offers MCMLE with starting values for estimators, as starting values close to the MLE is known to help reduce iteration rounds and avoid convergence failures for MCMC algorithms. Krivitsky (2017) found that MPLE typically outperformed CD as a seeding method for MCMLE in the binary ERGM regime; but since MPLE has not yet been implemented for Valued ERGMs, CD currently serves as the default seeding method for MCMLE in the `statenet` package for valued graphs. Here, we evaluate CD both as a standalone method and a seeding method, in comparison with MPLE, for MCMLE.

2.3.3 Maximum Pseudo-Likelihood Estimation

Although maximum pseudo-likelihood estimation (MPLE) has not to our knowledge been studied or implemented for count-valued ERGMs, it is an otherwise well-known technique (being the first practical method for general ERGM estimation (Strauss and Ikeda, 1990)). MPLE optimizes the product of the conditional likelihoods of each edge variable (the eponymous pseudo-likelihood (Besag, 1974)). In the unvalued case, this reduces to a logistic regression problem, allowing the MPLE to be obtained using standard regression algorithms (a fact that was instrumental in its early adoption, see e.g. Anderson et al., 1999). The MPLE is known to be consistent in some asymptotic scenarios (Hyvärinen, 2006; Strauss and Ikeda, 1990). For finite scenarios, in the special case of edgewise independent ERGMs, the MPLE coincides with the MLE; this ceases to be true for dependence models, though the MPLE is generally close enough to the MLE to be used as a standard method for initializing MCMLE estimators, and its first-order performance on large networks can be very good (An, 2016; Schmid and Desmarais, 2017). Because it does not fully account for interactions among edge variables, the pseudo-likelihood function tends to be excessively concentrated, leading to poor calibration of standard error estimates (as shown in binary ERGMs: Lubbers and Snijders, 2007; van Duijn et al., 2009). However, MPLE computation can be quite efficient, further aided by the fact that (1) the pseudo-likelihood itself can be approximated by subsampling edge variables, rather than computing on all of them, and (2) the calculations in question are embarrassingly parallel, making it possible to greatly reduce wall-clock time on multi-core CPUs.

As noted above, MPLE computation in the count-data context is more complex than in the binary case, and to our knowledge it has not been previously studied for count-valued ERGMs with dyadic dependence. We thus consider it here in greater detail. As in the binary case, the MPLE is defined by

$$\hat{\theta}_{\text{MPLE}} = \arg\max_{\theta} \prod_{(i,j)\in\mathbb{Y}} \Pr(Y_{ij} = y_{ij} | Y_{ij}^c = y_{ij}^c, \theta, X), \tag{2.5}$$

where the conditional probabilities in question are given by Eq. 2.3 and 2.4. Per Eq. 2.4, these latter conditionals depend upon a sum over the possible edge states of products of two factors: one involves the ratio of the reference measure at the observed edge value versus its alternative values, and the other involves the exponentiated difference in sufficient statistics between the observed network and the same network with the focal edge taking on alternative values. For the former, we observe that (in the case of the Poissonian reference), we have

$$\begin{aligned} \frac{h(\ell \cup y_{ij}^c)}{h(y_{ij} \cup y_{ij}^c)} &= \frac{(\ell!)^{-1} \prod_{(k,l)\in\mathbb{Y}\setminus(i,j)} (y_{kl}!)^{-1}}{(y_{ij}!)^{-1} \prod_{(k,l)\in\mathbb{Y}\setminus(i,j)} (y_{kl}!)^{-1}} \\ &= \frac{y_{ij}!}{\ell!}, \end{aligned} \tag{2.6}$$

while the latter is simply

$$\exp\left[\theta^T \Delta_{ij}(y,\ell)\right],$$

where Δ is the "generalized" changescore

$$\Delta_{ij}(y,\ell) = g(\ell \cup y^c_{ij}, X) - g(y_{ij} \cup y^c_{ij}, X).$$

There is not, in general, a simple form for the sum of these terms over all ℓ. However, we observe that the ratio of Eq. 2.6 falls very rapidly (as roughly $\ell^{-\ell}$) for $\ell \gg y_{ij}$, and it is hence possible in practice to approximate the infinite sum by truncation. More generally, we employ several techniques for improving computational performance, as described in the following subsections.

Pre-caching of ratios and differences

We note that neither the ratio of reference measures nor the changescores depend upon θ. Considerable computational savings can hence be had by pre-computing the ratios of Eq. 2.6 and the Δ values for the necessary range of ℓ values on each edge. This carries a storage cost that scales with the product of the ℓ range and the number of edge variables used, but avoids frequent recalculation of these (expensive) quantities on each pseudo-likelihood evaluation.

Edge sum truncation and/or coarsening

Per Eq. 2.3, the conditional log-likelihood of each edge variable involves as sum over $\ell \in 0, \ldots, \infty$. As noted above, we may approximate this sum by instead evaluating it over $\ell \in 0, \ldots, \ell_{\max}$, where $\ell_{\max}$ is large enough to be dominated by the decline in $y_{ij}!/\ell!$. Where the marginal distributions of each edge variable can be approximated as roughly Poissonian, choosing $\ell_{\max} = \lambda \max_{(i,j)\in\mathbb{Y}} y_{ij}$ with e.g. $\lambda \approx 4$ is an extremely conservative approach. (This is based on the observation that a Poisson random variable with expectation $z \geq 2$ has a 99.9% quantile for $4z$, assuming conservatively the expectation is the maximum observed value). λ can be further reduced as its max observed value increases (because the quantile of λz grows with z).

Truncation using the above method is adequate for small networks, or networks with low edge counts. However, when edge counts become extremely large, considerable computational effort may be wasted in computing conditional probabilities for small ℓ values when the observed value is large, or for large ℓ when the observed value is small. Using the same Poissonian approximation, we may further improve performance by working with the edgewise doubly-truncated sum over $\ell \in \ell^{ij}_{\min}, \ldots, \ell^{ij}_{\max}$, with $\ell^{ij}_{\min} = \max[0, y_{ij} - 4\lambda\sqrt{y_{ij}}]$ and $\ell^{ij}_{\max} = y_{ij} + 4\lambda\sqrt{y_{ij}}$. Because it is common to have network effects that can strongly suppress edges, however, we also recommend retaining some very small edge values as a buffer. Valued social networks usually have right-skewed distributions of edge values, so adding a few small edge values can also effectively cover the empirical distribution without significant increase in computation load. Our code by default retains integers from 0 to 5, although we strongly encourage extending coverage to the closest integer of sample mean of y, $[\mu]$ when feasible. This also defines the support of edges whose observed value is zero. The approach then leads to sums over ℓ values of the form $\ell \in \{0, \ldots, [\mu]\} \cup \{\ell^{ij}_{\min}, \ldots, \ell^{ij}_{\max}\}$. This usually retains $\mathcal{O}(\sqrt{y_{ij}})$ terms per sampled edge, which is often a substantial savings as y_{ij} values become large.

When dealing with extremely large counts, storing and computing even $\sqrt{y}$ terms can become prohibitive (particularly if many edge variables are needed for adequate statistical power). In such cases, a coarsened approximation to the sum is another option. To coarsen, we select k evenly spaced values from $\ell^{ij}_{\min}, \ldots, \ell^{ij}_{\max}$, and compute the associated contributions to the edge sum only for these terms. We also include, however, the gap (in terms of the number of "skipped" ℓ values) between subsequent calculated terms, and weight each computed term by the number of elements in the gap; this is equivalent to approximating the sum via a step function, with knots at the computed ℓ values. Our experience with this method has been promising, although problems can ensue if the sum becomes heavily concentrated on a range of terms that lie within adjacent knots. We thus do not employ this technique in this chapter, although we offer it as a promising target for future research. Of course, other approximation methods are also possible (e.g., integral approximations), and may be useful in the large-count regime.

Edge variable sampling

Although the exact calculation of the pseudo-likelihood is at least $\mathcal{O}(RN^2)$, the log pseudo-likelihood can easily be approximated by random sampling of edge variables; this reduces both storage and computational cost. As shown by our experiments, subsampled MPLE can yield high-quality estimates with less time consumed. Our implementation offers different sampling schemes such as uniform random sampling, as well as weighted (i.e., importance) sampling schemes analogous to the "Tie-No Tie" proposal method frequently used in ERGM MCMC (Morris et al., 2008). The two schemes are almost identical in our study case because the binary density is close to 0.5, and we use the random sampling scheme in this chapter for simplicity.

Parallel evaluation of conditional log-likelihoods

Because the log of the pseudo-likelihood is linearly separable, its calculation is an embarrassingly parallel problem. In practice, we divide sampled edge variables into batches, and calculate their conditional log-likelihoods independently on different cores. This leads to wall-clock time reductions, as the pseudo-likelihood calculation time scales with the inverse of the number of available cores. This (combined with edge variable sampling) can make the MPLE an attractive choice for very large valued networks, especially when many cores are available.

Taken together, the above computational techniques allow the MPLE to be used even for very large networks with highly dispersed counts (although not all of them are needed when counts are less variable, or on smaller networks). As we show, valued MPLE is very fast, and the resulting estimator can have low bias and high accuracy for valued networks; it offers high-quality calibration of uncertainty when the edge variance is large, but is prone to overconfidence for dependence terms (i.e., underestimation of the standard error) and conservative for nondependence terms when the edge variance is small.

2.4 Study design

We evaluate the above estimation techniques via a parameter recovery study, in which we generate networks from a realistic generative model based on an initial fit to real-world social networks, estimate models to the simulated draws using each respective technique, and then examine the properties of the resulting estimators. Our generative model was created by fitting a Valued ERGM to an empirical case (see below) using MCMLE; we then obtained 500 high-quality draws from the fitted ERGM using MCMC. For each draw, we obtained point and standard error estimates from each of the three study methods (MCMLE, CD, MPLE), evaluating the results with respect to wall-clock estimation time, bias, variance of the estimator, overall accuracy, and calibration (accuracy of estimated standard errors and confidence coverage). All modeling and analysis was performed using `statnet` (Handcock et al., 2008), specifically using the `ergm` 4.2.1 (Hunter et al., 2008b; Krivitsky et al., 2022), `ergm.count` 4.1.1 (Krivitsky et al., 2012), and `sna` 2.6 (Butts, 2008) libraries. Our MPLE implementation also made use of the `Rcpp` library (Eddelbuettel and Balamuta, 2018). The following subsections detail the data and model used, the setup of the estimation procedures, and the performance metrics.

2.4.1 Case Study and Model Definition

To examine the performance of estimation methods for data with different network sizes and edge value variances, we construct the following study cases. They are based on real-world datasets of migration flows between counties in two U.S. states (New Mexico and North Carolina), collected and released by the American Community Survey (U.S. Census Bureau, 2019). The New Mexico data consists of 33 nodes and the North Carolina data 100 (henceforth the "small" network and the "large" network, respectively). For the New Mexico data, we generate two networks with different edge value distributions. The large-variance case uses the count of migrants between each directed county pair as the edge value, which ranges in integers from 0 to 3,862 with standard deviation 201. For comparison, the edge value of the small-variance case takes the natural logarithm of migrant count (plus one), rounded to the nearest integer. Its edge value ranges from 0 to 8, with standard deviation 2. Ideally, we would generate large-variance and small-variance cases for the large network as well. Unfortunately, the large-variance large-network case turned out to be prohibitively computationally expensive for a simulation study comparing all standard methods, and we hence just include the small-variance large network case; the edge value is generated by the same manner discussed above, and the distribution is similar to the small network case, ranging from 0 to 9 with standard deviation 2. Table 2.1 displays the descriptive statistics of the three study cases.

Our ground-truth model is created by fitting a count-valued ERGM to the above networks. The sufficient statistics include a Sum term (intercept, the summation of all edge values), a Nonzero term (the count of nonzero edges), three exogenous covariates, and two depen-

Table 2.1: Network Descriptive Statistics of the Studied Cases

	Network	Binary	Edge value			
	size	density	min	max	mean	std. dev.
Large-variance small network	33	0.50	0	3,862	46.15	200.54
Small-variance small network	33	0.50	0	8	1.64	1.95
Small-variance large network	100	0.41	0	9	1.32	1.84

dence terms. The exogenous covariates are the population sizes of the sending and receiving counties (called Nodeocov and Nodeicov respectively), and the distance between counties (Edgecov) (all on natural log scale), using 2010 Census data (National Bureau of Economic Research, 2016; U.S. Census Bureau, 2011b). The first dependence term, mutual, measures the reciprocity of the network, defined as

$$g_{\mathrm{m}} = \sum_{(i,j)\in\mathbb{Y}} \min\{y_{ij}, y_{ji}\}$$

The second dependence term, flow, is adapted from empirical analyses of inter-county migration networks in the United States (Huang and Butts, 2024a, 2023). It calculates the summation of the volumetric flow of each node, which is the minimum of total inflow and total outflow for a node. It is a count-valued version of two-paths or mixed-two-star terms for binary networks (Morris et al., 2008). Formally,

$$g_f = \sum_{i\in\mathbb{V}} min\{\sum_{j\in\mathbb{V}\setminus i} y_{ij}, \sum_{k\in\mathbb{V}\setminus i} y_{ki}\}$$

where $\mathbb{V}$ is the vertex set. This model encompasses a diversity of different sufficient statistics commonly used in valued ERGMs, including graph-level baseline statistics (Sum, Nonzero), covariate effects (Nodeicov, Nodeocov, Edgecov), and dependence terms at the dyadic level (Mutual) and the triadic level (Flow). Although our aim here is to produce a deliberately simple model for purposes of evaluation (as opposed to a substantively detailed model of migration), our choice of statistics was informed by prior work on migration, and previous empirical analyses on migration-flow networks in particular (Boyle et al., 2014; Huang and Butts, 2024a, 2023; Windzio, 2018; Zipf, 1946).

2.4.2 Methods under Evaluation

As described above, we estimate parameters from the simulated network draws using three procedures: Contrastive Divergence (CD), Maximum Pseudo-Likelihood Estimation (MPLE), and Monte Carlo Maximum Likelihood Estimation (MCMLE). Estimation for each method was performed as follows.

For CD, we use the default settings of the `ergm.count` package, performing 8 Metropolis-Hastings steps, raising one proposal in each step. We also tried using more steps and/or more proposals within each step for CD. As is shown in the Appendix, its estimation bias and calibration usually do not improve systematically as we increase the tuning parameters, and when it does, it fails to match other methods in comparable time. Therefore, we keep the most time efficient setting in our comparison.

For MPLE, we implement the procedure described in Section 2.3.3. To examine the performance of MPLE with various sample sizes, we consider three subsampled MPLE in each study case: the fast, the mid, and the full, which corresponds to uniformly sampling 50%, 75% and 100% of the edge variables in random, respectively; this corresponds to 528, 792, 1056 dyads for small networks, and 4950, 7425, 9900 dyads for the large network. We also consider the impact of multiple cores on execution time, calculating the wall-clock time for models estimated using 1,4, and 20 cores, respectively. In terms of edge support truncation, for small-variance cases, we use a uniform non-edgewise truncation; the support covers integers from 0 to λ times the max edge value of the network, where $\lambda = 4$ for the small network and $\lambda = 1.5$ for the large network as the latter has more information with more edge variables. Edgewise support truncation becomes a powerful tool to reduce computation time for large-variance networks; we set $\lambda = 4$ for the doubly-truncated edgewise support (see the second paragraph of Section 2.3.3). We also coerce the support to include integers from 0-value to the 80% quantile of the edge distribution for every edge variable, whose upper bound ranges from 43 to 48 and is typically close to the mean of the edge distribution. This is a conservative scheme for an edgewise support truncation with wide intervals and without coarsening, but is fast enough to have one order of magnitude less wall-clock time than MCMLE.

For MCMLE, we use the stochastic approximation method in `ergm.count`, a workhorse method that is also implemented and served as the default method in PNET (Wang et al. 2009). We made a few adjustments from the default setting to improve its performance based on the data structure of the study case and our exploratory experiments. First, we set the proposal distribution of Metropolis-Hastings algorithm in MCMC to be random, where every dyad has equal chance to be toggled. By default, the proposal distribution in `statnet` favors toggling nonzero edges than nulls, with the rationale that social networks tend to be sparse (for binary networks); we revoke this penalty towards nulls since the binary density of the network is high (0.4-0.5) in the study cases (removing the need for biased proposals), and the random proposal reduces the computational time of the MCMLE. Second, by default, the MCMC thinning interval is 1024 and the sample size of network statistics in each distribution returned by the algorithm is also 1024.[2] This setting is sufficient for the small-variance small network case, but for the other two cases, we increase these two parameters to be ten thousand; this helps with convergence and further increasing those parameters no longer brings performance gain based on our experiments. We evaluate the use of both CD and MPLE as seeding methods for MCMLE, referring to them as CD-MCMLE and MPLE-MCMLE, respectively. We employ default settings for CD, given that longer chains do not consistently enhance performance (see Appendix). For the small-variance small network case, we use the “mid” MPLE for MCMLE seeding, but for the large-variance and the

---

[2] We follow the default of `statnet` that sets MCMC burnin as 16 times the length of the thinning interval.

large network cases, we use the "fast" MPLE since the fast MPLE yields good-enough point estimations. It turns out that CD-MCMLE fails to converge for some of the large-variance cases; we detect this by examining the MCMC diagnostics plot, and reruning those cases until convergence (details in Table 2.2 in Section 2.5). We also turn off the bridge sampler that calculates the log likelihood to reduce the MCMLE computational time, since this is not involved in estimation.

All models were fit on a 44-core server, with 256GB RAM. The processors are dual Intel Xeon E5-2699 2.2GHz CPUs (22 cores/CPU). Estimation using R 4.1.1 was performed on Ubuntu 20.04.1. All procedures reported are based on a single core, except for the multi-core MPLE conditions.

2.4.3 Evaluation Criteria

Since the methods of interest involve speed/quality trade-offs, it is necessary to evaluate these two dimensions simultaneously. To evaluate computational cost, we compute the wall-clock time for each method, as mean seconds needed to fit the target model to a simulated network. Since the speed of MPLE is dependent upon the number of parallel processes, we repeat the process using 1, 4, and 20 cores, respectively.

To evaluate estimation quality, we consider the bias, variance, overall accuracy, and calibration of each estimator using the following metrics.

We first compute the absolute relative bias (ARB) of estimators for each coefficient, using

$$ARB = \left| \frac{1}{m} \sum_{i=1}^{m} \frac{\hat{\theta}_i - \theta}{\theta} \right|$$

taking the average across m experiments, where $\hat{\theta}$ is the estimator and θ is the true value from the model that simulated the networks. The smaller the ARB, the less bias is introduced by the estimation procedure.

We also compute the variability for each estimator, via the (true) standard error of the estimated coefficient. Formally,

$$SE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(\hat{\theta}_i - \frac{1}{m}\sum_{j=1}^{m}\hat{\theta}_j)^2}$$

The smaller the variability, the more efficient and more precise the estimator is.

While bias and variance are each important, we are also interested in the total accuracy of the estimator (the extent to which it deviates, on average, from the true value). We measure this via the root-mean-square error (RMSE) i.e.

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(\hat{\theta}_i - \theta)^2}$$

The smaller the RMSE, the more accurate the estimator is on average.

Finally, we consider how well calibrated each estimator is, in terms of the associated estimates of uncertainty. To evaluate the bias in our second moment estimate, we compare the real standard error se and the estimated standard error $\hat{se}$ using

$$Calibration = \log\left[\frac{1}{m}\sum_{i=1}^{m}\frac{\hat{se}_i}{se}\right]$$

A positive number suggests the method is conservative, while a negative number suggests the method is overconfident. We also examine confidence coverage, specifically the proportion of cases in which the nominal 95% confidence interval (CI) for each parameter actually covers the true coefficient. Specifically, the coverage rate is computed by

$$Coverage = \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}\{\theta \in [\hat{\theta}_i - z \cdot \hat{se}_i, \hat{\theta}_i + z \cdot \hat{se}_i]\}$$

where $z = 1.96$ for 95% CI. The closer to 95% the coverage is, the better calibrated the estimate is. Coverage rates above 95% suggest that the method is conservative, while coverage rates less than 95% suggest that the method is overconfident.

2.5 Results

2.5.1 The Small-variance Small Network Case

Starting from the simplest case of the small-variance small network, we display the performance of each method in Figures 2.1 and 2.2. Panel A in Figure 2.1 shows the absolute relative bias (ARB) of the coefficient estimates. It shows that all methods produce very small numerical biases, 3% or less across all covariates and methods. CD and fast MPLE introduces larger biases; but as the sample size of MPLE increases, its bias reduces and gets close to that of MCMLE, seeded by either CD or MPLE. This finding is consistent with research on binary ERGMs finding that the MPLE introduces little bias in parameter estimation (van Duijn et al., 2009; Schmid and Desmarais, 2017). The lack of appreciable bias is an encouraging sign, suggesting that point estimation of valued ERGMs for small-variance small network is easy to acquire using whichever method we tested.

We also evaluate the (im)precision or variability of estimation (sometimes called efficiency),

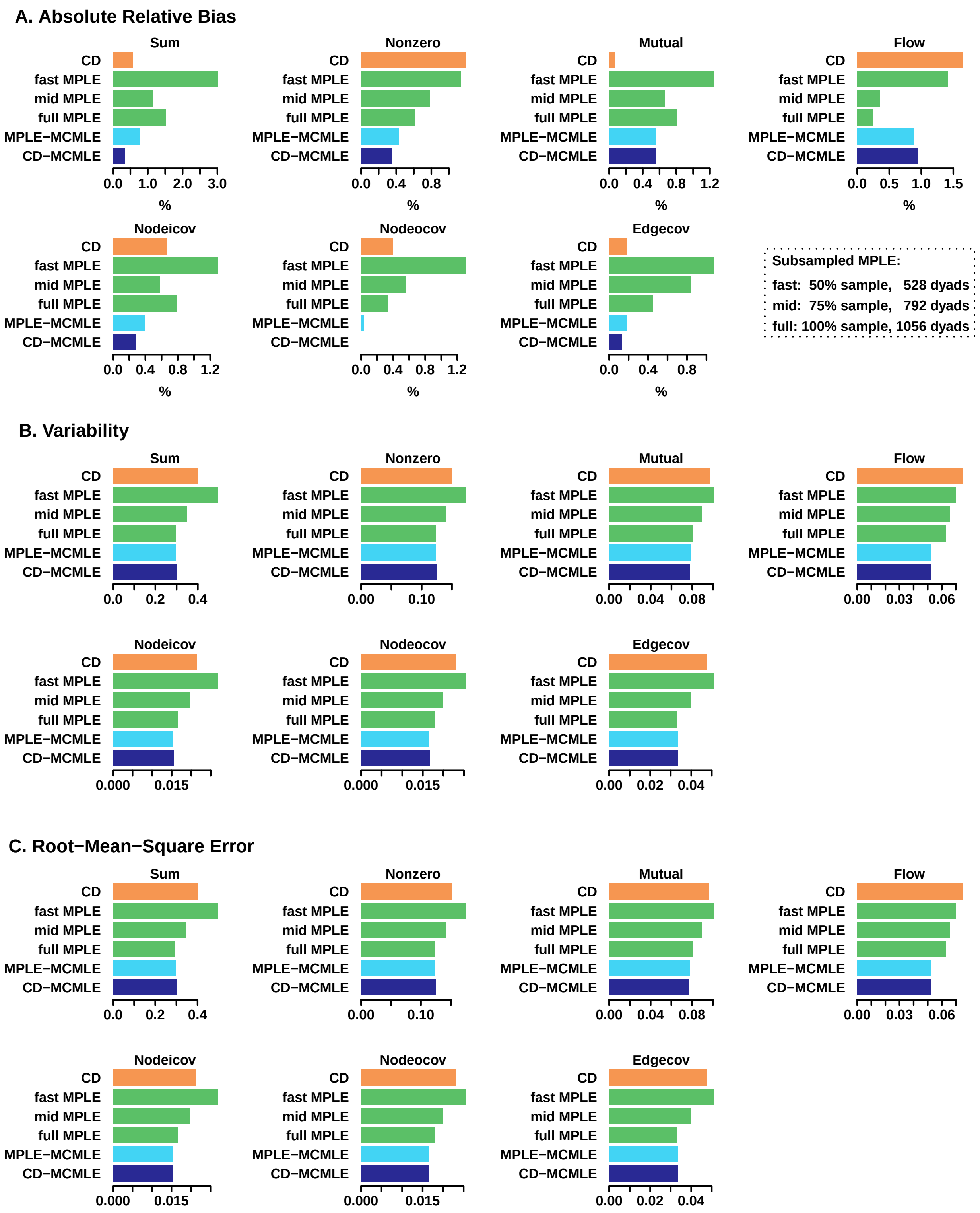


Figure 2.1: Bias, variability, and RMSE of small-variance small network

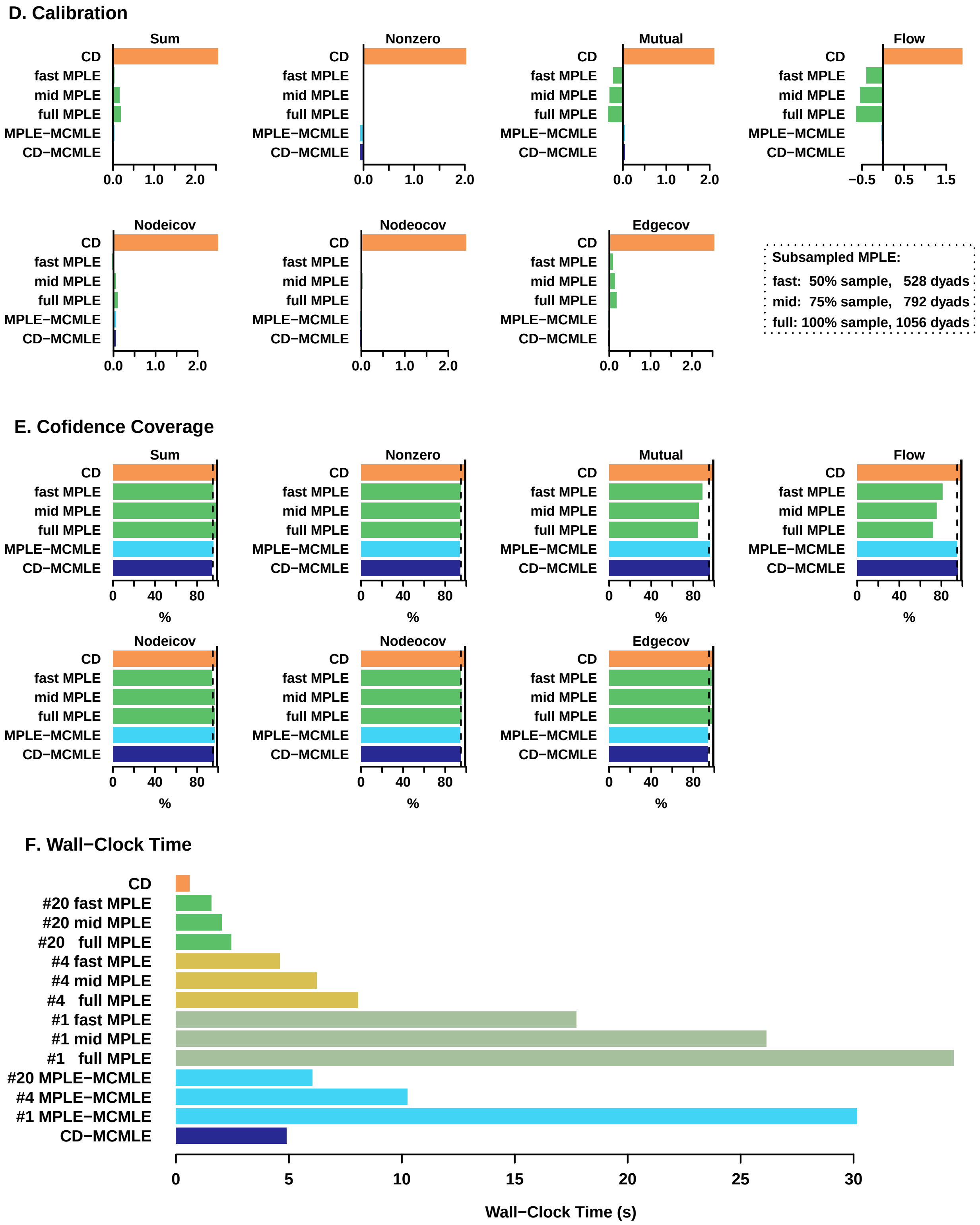


Figure 2.2: Calibration, confidence coverage, and wall-clock time of small-variance small network

i.e. the true standard error of each estimator. Panel B in Figure 2.1 shows that the variability of the MPLE decreases with more edges sampled, and that full MPLE and MCMLE are the most efficient methods. In general, variations of estimators using all methods are close to each other, suggesting that they have similar efficiency.

We then evaluate the total accuracy of estimators using the root-mean-square error (RMSE). A more holistic metric, the accuracy measurement combines bias and variation of estimation evaluated above, and smaller RMSE is preferred. Panel C of Figure 2.1 displays RMSE scores, whose distribution is almost identical to the variability scores in Panel B. The similarity between RMSE and variability reveals that biases contribute very little to the total RMSE, with accuracy being dominated by the performance in variability. Methods with good variability score thus also have decent accuracy. Full MPLE and MCMLE has the smallest RMSE, though RMSEs for all methods under evaluation show only mild differences.

Besides performance in coefficient estimation, performance in estimating uncertainties is also evaluated, shown in the first two panels in Figure 2.2. Panel D displays calibration of each method. An indicator of the bias in standard error estimation, a positive calibration score suggests overestimation of the uncertainty, and a negative calibration suggests underestimation. Noticeably, CD overestimates standard errors for all covariates by a large margin; CD's calibration scores are all above 1.9, indicating that the estimated standard error is more than 6.6 (i.e., $e^{1.9}$) times its true value. We experimented with different tuning parameters for CD, but could not find settings with both improved calibration and reasonable execution time (see Table 2.A2 in Appendix). In summary, CD is too conservative to offer useful uncertainty estimations of covariates for small-variance small networks.

For MPLE, we find that it underestimates the uncertainties for the dependence terms (mutual and flow), but overestimates uncertainties for non-dependence terms, though the degree of inflation is very small. Previous simulation studies found similar patterns for MPLE on binary ERGMs (Lubbers and Snijders, 2007; van Duijn et al., 2009). Interestingly, the bias in standard error estimation increases with the sample size for MPLE. This is in part because the bias of estimation in statistical uncertainty is trivial when the numerical uncertainty is the main source of uncertainty for the MPLE (as is the case with small sample sizes); as the numerical uncertainties decrease with more edges sampled, the bias in statistical uncertainty becomes non-negligible. The calibration of MPLE is not as far off as that of CD, but its underestimation of standard errors is noticeable for dependence terms in the small-variance small network case.

While CD and the MPLE show varying degrees of error in standard deviation, both CD-MCMLE and MPLE-MCMLE have almost perfect calibration, suggesting that MCMLE is the best method for standard error estimation of small-variance small network.

Another metric that considers uncertainty estimation, confidence coverage is the proportion of model fits in which the true value of a given coefficient is covered by the estimated 95% confidence interval (CI), as is shown in Panel E of Figure 2.2, where the dotted line is the 95% reference line and the solid line represents 100%. The figure tells a similar story to Panel D's calibration score, because confidence coverage performance is largely determined by performance in calibration of uncertainty when the bias of coefficient estimations is small. The figures show that CD overestimates standard errors so much that its CIs always cover the true value, making them conservative but uninformative. The MPLE's CIs cover the true values more than 95% for non-dependence terms, but under-cover the true values for the dependence term, with this deviation becoming larger as sample size increases. On the other hand, both CD-MCMLE and MPLE-MCMLE have coverage rates that are extremely close to 95%, showing its characteristic calibration advantage for small-variance small network of valued ERGMs estimation.

Lastly, Panel F of Figure 2.2 displays the wall-clock time of each method. As expected, the wall-clock time for computing the MPLE can be greatly compressed by using a sample of edges to approximate the joint pseudo-likelihood function, or by using multiple cores to calculate conditional likelihoods. The fastest methods are CD followed by MPLE using 20 processors. Overall, the wall-clock time for different methods is short and varies modestly for this simple computation case, costing half a minute at most. Interestingly, while MCMLE is commonly believed to be a slow method, it is very fast in this simple case.

In summary, for small-variance small network data, all methods offers accurate and minimally-biased point estimates. CD offers uninformatively conservative uncertainty estimates, and MPLE's uncertainty estimation for dependence terms is noticeably overconfident. All methods are reasonably fast in this regime. With great performance in all metrics, MCMLE is an ideal method for valued ERGMs estimation for small-variance small network data.

2.5.2 The Small-variance Large Network Case

The small-variance large network case has similar edge value distribution to the previous small network case, but its network size is 3 times bigger, meaning that its dyad count is 9.4 times the count of the previous case. Comparing these two cases offers insights about the influence of network size on estimation performance for each method.

Figure 2.3 shows that, again, all methods introduce very little bias. One noticeable difference is that the bias of subsampled MPLE gets smaller in comparison with other methods. This suggests that the absolute number of edge variables sampled influences the performance of subsampled MPLE; for small networks, one needs to sample a larger proportion of edges, while for large networks, the percentage can be lower. This means that for the large network case, the fastest MPLE is already less biased than CD (whose bias from a larger tuning parameter setting gets outperformed by following up CD with MCMLE, see Table 2.A1 in Appendix); this translates to the less biased performance of MPLE-seeded MCMLE, compared to the CD-seeded MCMLE as the figure shows. Panels B and C in Figure 2.3 reveal almost identical patterns in variability and RMSE compared to the small-variance

Figure 2.3: Bias, variability, and RMSE for small-variance large network

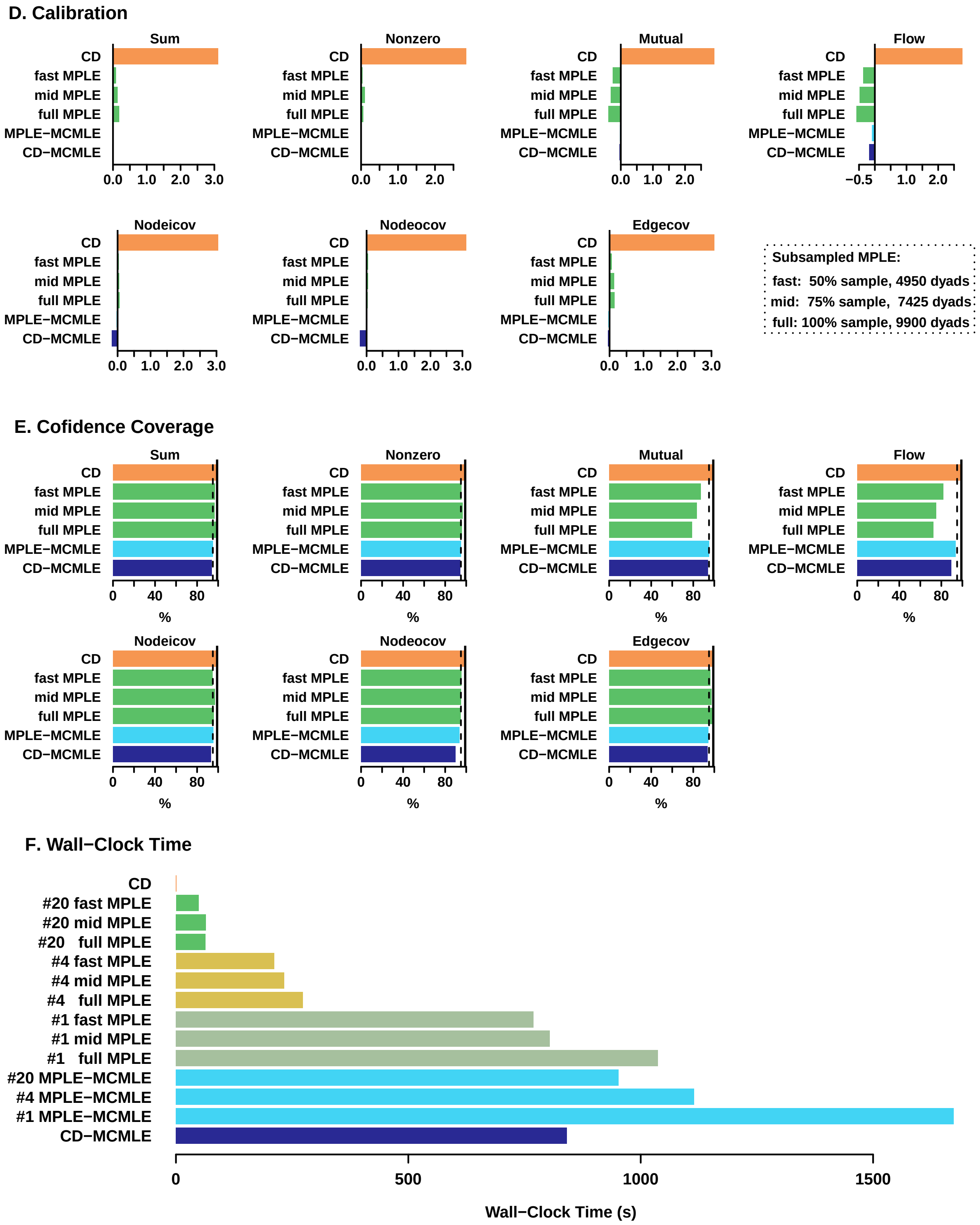


Figure 2.4: Calibration, confidence coverage, and wall-clock time for small-variance large network

small network case. Variability decreases as MPLE's sample size increases, and all methods share similar variability. The RMSE distributions resembles those of the variability as the bias of all methods are largely ignore-able.

Comparing Panels D and E in Figure 2.4 with those in the previous Figure 2.2 suggests that patterns of uncertainty estimation performance across methods are generally invariant to network size. We again find that CD greatly overestimates uncertainty, leading to over-coverage of the confidence intervals; MPLE underestimates the uncertainty for dependence terms and the confidence intervals under-covers their true values. MCMLE offers great uncertainty estimation again, although, on close inspection, CD-MCMLE shows a (very) small tendency towards overconfidence that the MPLE-MCMLE lacks.

Panel F in Figure 2.4 shows that the wall-clock time of MPLE and MCMLE scales with the network size, while that does not apply to CD. Subsampling and use of multiple cores effectively reduce the computational time of MPLE; this advantage becomes larger with network size, since larger networks require a greater share of computing time to be used for change score calculation. Since changescore calculations are embarrassingly parallelizable, gains from multi-core calculations grow accordingly in this regime. Overall, this scenario shows clear superiority of CD for computational time, followed by MPLE using subsampling and multi-core strategies. MCMLE becomes quite slow here (with mean times between ≈ 13 minutes and roughly half an hour), making speed a potentially important consideration.

To recap, comparing the two small-variance cases with small and large network sizes suggests that the estimation quality of each method is largely invariant to the network size. The larger number of edge variables for the large network means that subsampled MPLE requires a smaller share of edge variables for good performance, and fast MPLE becomes a less biased and a better seeding method than CD for MCMLE (though the difference is small in the study cases). All methods have good first-order performance, though MCMLE is clearly superior for calibration (with CD being unacceptably poor). The major performance difference coming from the network size is that the superiority in computational efficiency for CD and multi-core subsampled MPLE becomes substantial when the network size grows.

2.5.3 The Large-variance Small Network Case

Given the consistency seen in the two small-variance cases, we might expect the large variance, small network case to behave similarly. However, we observe very different results in the small network case when the variance of edge values becomes large. First of all, we observe that CD-MCMLE simply fails to converge for some of the simulated networks, as reflected by their MCMC diagnostics plots; here we follow the common procedure of rerunning them until convergence, though failure to attend to diagnostics could lead to problems in casual use. Table 2.2 summarizes the number of rounds CD-MCMLE went through before seeing convergence. Overall, it took 1.22 rounds on average for CD-MCMLE to converge (with a few cases taking more than five). The following results are based on their final (converged) rounds, as estimators from the failed rounds were very far from the true values.

Table 2.2: Number of rounds for CD-MCMLE before convergence

N of Rounds	1	2	3	4	5	6	7
Count	434	38	17	9	0	1	1
Percentage (%)	86.8	7.6	3.4	1.8	0	0.2	0.2

Panel A in Figure 2.5 shows that CD introduces relatively larger biases for the large-variance case, 5.2% and 3.8% for the nonzero and the flow terms, respectively. Although those biases are arguably not huge, they lead to failures in convergence for CD-MCMLE. On the contrary, MPLE introduces very little bias in its estimates, the largest bias of 1.8% coming from the fast MPLE for the flow term. That makes it an excellent seeding method, and indeed all MPLE-MCMLE models converged in their first attempt. This signifies that, similar to the binary ERGM scenario, MCMLE for valued ERGMs is sensitive to the seeding quality, and small improvements in biases of the seeding methods can make a difference. Another notable

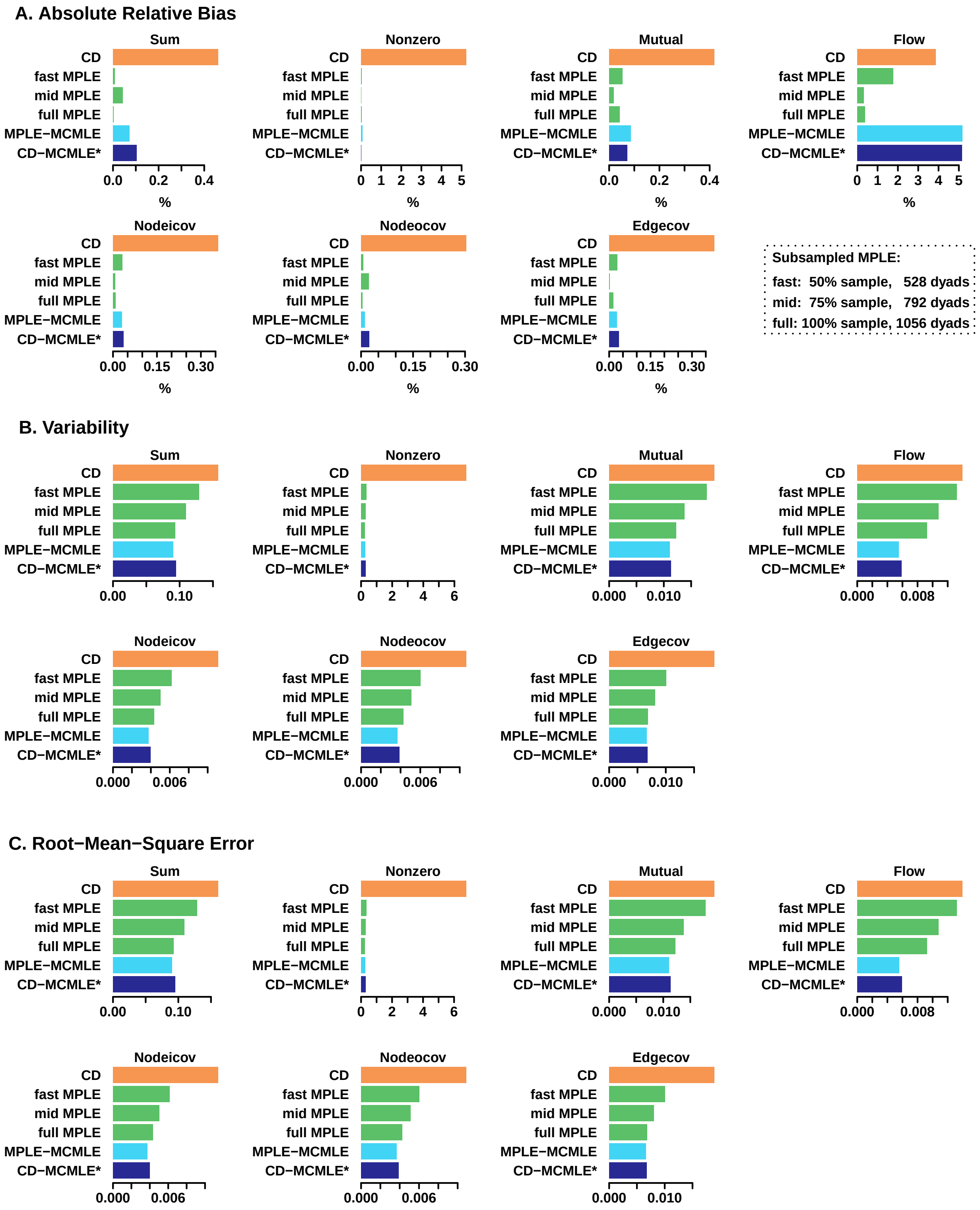


Figure 2.5: Bias, variability, and RMSE of large-variance small network

Note: CD-MCMLE* results are from their final rounds with convergence.

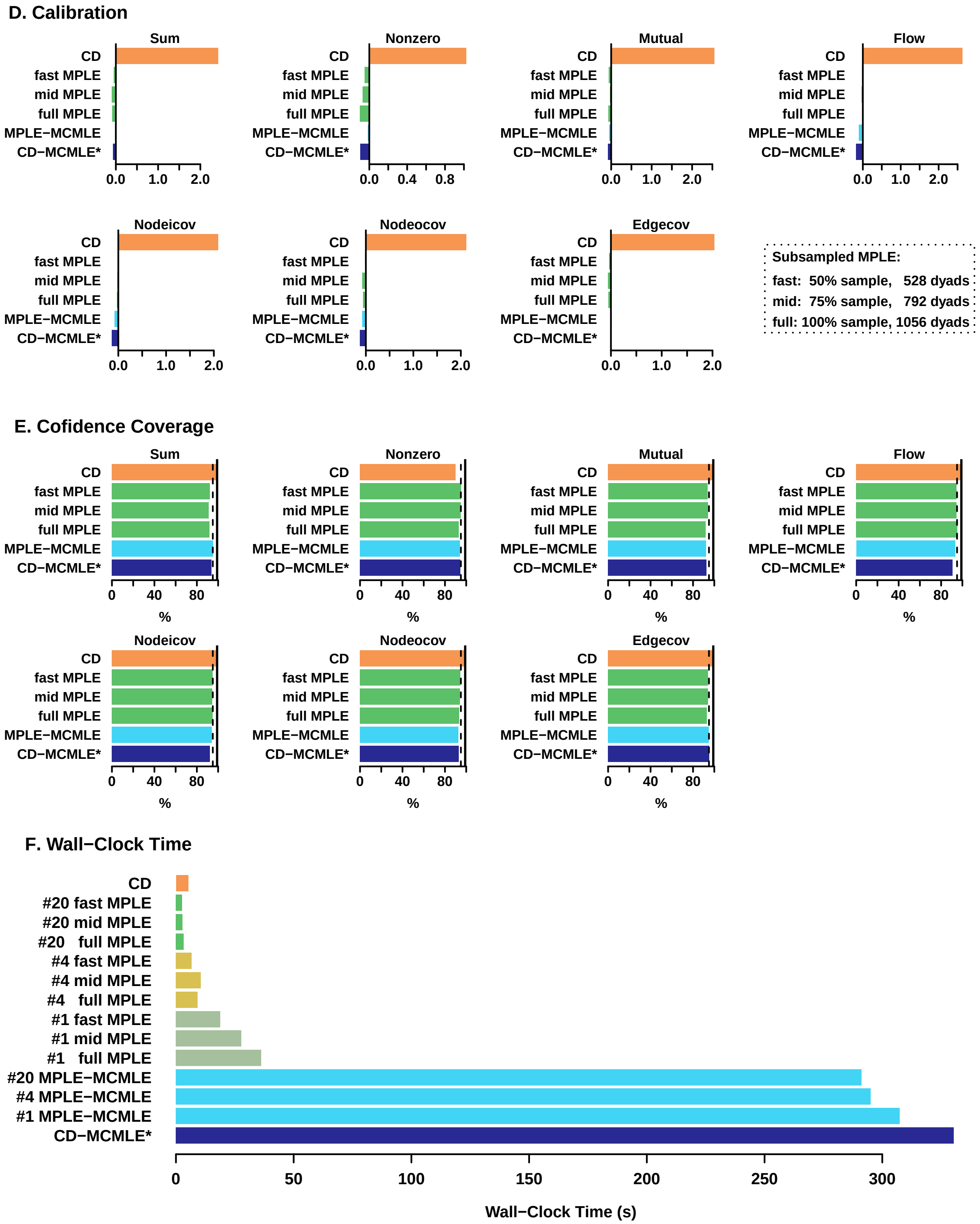


Figure 2.6: Calibration, confidence coverage, and wall-clock time of large-variance small network

Note: CD-MCMLE* results are from their final rounds with convergence. Its wall-clock time is simple summation of all rounds of computation.

feature is that MPLE actually outperforms MCMLE in the bias metric, with the latter having a bias of about 5% for the triadic dependence term (flow).

Panel B in Figure 2.5 shows that CD generally has larger variability than other methods, especially for the nonzero term. MPLE's variability decreases with more edge variables sampled, and gets close to MCMLE when all edges are utilized. Panel C in Figure 2.5 demonstrates that, again, when biases are generally small, the accuracy metric resembles that of the variability.

Panel D in Figure 2.6 shows that CD substantially overestimates the standard errors, so overly conservative uncertainty estimation is a consistent behavior for CD across all network size and edge variance structures studied. By contrast, MPLE no longer suffers from calibration difficulties for the large-variance case, producing high-quality standard error estimates that match the performance of MCMLE, seeded by either CD or MPLE. Note that for the nonzero term, CD did not return an estimate for 22 of the 500 simulated networks, meaning that the calibration of CD for the nonzero term could actually be worse; we did not rerun them, as CD's uncertainty estimates are not useful even when the algorithm did converge. Panel E in Figure 2.6 reveals that CD's confidence intervals have the over-coverage issue for every covariate except the nonzero term, which suffers from a larger first-order bias. MPLE and MCMLE, both with small bias in point estimation and uncertainty estimation, offer confidence coverage very close to the 95% benchmark.

Lastly, Panel F in Figure 2.6 shows that for the large-variance network, MCMLE is an order of magnitude slower than CD and MPLE. Subsampling and use of multiple cores are still effective ways of reducing computational time for MPLE, and MPLE with 20 processors becomes even faster than CD. We should also note that the time reported here for CD-MCMLE is the simple summation of their all rounds of wall-clock time. In reality, users need to spend more time digging into the diagnostics of MCMC after each round, and this makes CD-MCMLE even more slower than running MPLE-MCMLE.

Overall, we observe that the variance of edge values makes a substantial difference in the behavior of these estimation methods. CD generates fair point estimates in a speedy manner (albeit less accurate than its peers), but its calibration is overly conservative to the point of being unusable. The larger biases of CD estimators sometimes prevents convergence of CD-MCMLE, making MPLE a better seeding method in this scenario. MCMLE, seeded by either CD or MPLE, once converged, offers high-quality estimates at correspondingly high cost. Strikingly, however, we find that in this case MPLE generates estimators that match MCMLE in all metrics and introduces even less bias. Considering that MPLE is also an order of magnitude faster than MCMLE, it is clearly the superior method for the large-variance case. If higher-quality estimates are needed, one can increase the sample size of MPLE or follow it up with MCMLE, though the former is a much faster option.

2.6 Discussion

Overall, this comparative simulation study reveals that the variance of the edge value makes a substantial difference in the performance of estimation methods for Valued ERGMs, while network size primarily impacts computational cost. For small-variance data, all methods perform very well in point estimation, while CD greatly overestimates the uncertainty, and MPLE underestimates uncertainties of dependence terms. MCMLE seeded by either CD or MPLE offers high-quality estimates under all metrics. Wall-clock time of the methods are close to each other for small-variance small networks, but the speed advantages over MCMLE gets larger for CD and subsampled/multi-core MPLE as the network size increases. For very large graphs (especially when large numbers of cores become available), the relative speed advantage of the MPLE can become substantial, and may be a reasonable consideration in method selection.

For large-variance data, CD fails to offer reliable uncertainty estimates; its relatively high bias compared to MPLE also makes CD-MCMLE more prone to convergence failure. Both MPLE and MCMLE (given convergence of the latter) perform well in all quality evaluations, but MPLE is an order of magnitude faster than MCMLE, and can be further sped up by subsampling and parallel computing. Again, we observe that the speed advantages of MPLE become larger as graph size and edge variance increase.

The results suggest that for small networks with low-variance edges, MCMLE continues to serve as the estimation method of choice: it delivers high-quality point estimates and excellent calibration while still being computationally accessible in this regime. As network size increases, MCMLE becomes increasingly cumbersome, as its computational time scales up faster than other methods. When MCMLE is too slow to perform well, CD and MPLE can serve as useful tools for tasks that only require first-order point estimation, such as exploratory analysis, prediction, or generating models for network simulation. The subsampling and multi-core features of MPLE offer useful tools for fast computation of models on large networks, and the numbers of edge variables that must be sampled for strong performance are an increasingly small fraction of the total dyad count as network size grows, further enhancing its computational advantages. Although MPLE calibration is certainly good enough to be useful (especially for independence terms), we clearly see the tendency towards overconfidence in dependence terms found in binary ERGM studies, and nominal confidence intervals for these terms are likely to be too small. Analyses relying on coverage for such terms should be regarded as heuristic (though the maximum extent of miscalibration seen here may provide some guidance with respect to the degree of error that could be present).

For networks with large edge variance, MPLE could be the go-to method, yielding accurate estimators with good calibration. Besides subsampling and parallel computing choices, our implementation of edgewise support truncation further offers MPLE an edge in speed without compromising its estimation quality (as edgewise and nonedgewise truncation of MPLE offers commensurate results in quality metrics for large variance data). This enables the use of MPLE to estimate Valued ERGMs that were previously blocked by computational barriers, such as high-dimensional models on large networks with high edge variance. It also offers a flexible framework to make trade-offs between estimation quality and computational time, by tuning the sample size of the edge variables and the structure of the edge support truncation. While MPLE has been seen as a sub-optimal choice for binary ERGMs (van Duijn et al., 2009), this comparative simulation study reveals an area where it can be the effective approach.

Our experiments also offer insights about MCMC-based estimation methods. This chapter shows that MCMLE for Valued ERGMs depends on high-quality starting values, especially for large-variance networks. Both CD and MPLE are useful tools for MCMLE seeding, but the relatively larger biases of CD for large-variance data makes CD-MCMLE more fragile. One potential reason for this fragility is the difficulty of CD in reproducing the dichotomized density of the networks (i.e. the proportion of nonzero edges). With larger ranges of edge values, the toggling of values between zero and one become more unlikely, and the difficulty of matching the target density increases. (This is, of course, a special case of zero-inflation, a common phenomenon in count data models beyond network settings.) This is reflected by the phenomenon that the sufficient statistic consistently observed to fail in CD-MCMLE was the nonzero edge count in the large-variance experiment. The issue can be worsened by the mismatch of the MCMC algorithm design and typical properties of valued networks. Compared to their binary counterparts, valued social networks are frequently denser (in the dichotomized density). However, MCMC algorithms in existing software are optimized for sparse, unvalued social networks. We observed improvement in computational time when switching from an MCMC proposal that favors toggling nonzero edges (so-called TNT, or more accurately "tie-random dyad") to a random proposal, which is the one that offers the highest likelihood of toggling empty edges among existing algorithms. To improve the performance of MCMC-based methods for Valued ERGMs, future research could consider experimenting with MCMC algorithms that pay more attention to the toggling between value zero and one, such as proposals that favors toggling zero-value edges.

As with any simulation study, one trades off the “realism” of performance on a realistic case against some degree of generality. Although we vary the network size and edge value variance to emulate different application settings, we cannot rule out the possibility that some methods studied here may perform better or worse under other conditions. We encourage future research using simulation studies based on different use cases and model specifications, including non-Poissonian reference measures. Given that we find that the edge value variance plays an important role in influencing the performance of Valued ERGM estimation, it would be of interest to experiment with more fine-grained classification of the scale of edge variance, in search of an empirical rule of thumb for when MCMLE or MPLE would be the better choice.

Our study also suggests the continuing relevance of the MPLE to ERGM methodology. Our implementation of MPLE for Valued ERGMs enables estimation for large-variance data in feasible time and with high-quality results. With good overall accuracy, high speed, and flexible tunability, MPLE would be an excellent general use estimation method for Valued ERGMs if its calibration could be improved for small-variance networks. Our findings suggest the value of work on methods that may help further improve calibration of MPLE in the count-valued case; such advances may build on methods that shown to help calibration for binary ERGMs, such as bootstrap resampling (Desmarais and Cranmer, 2012a; Schmid and Desmarais, 2017) and regularization (van Duijn et al., 2009).

Lastly, we should emphasize that it is important to perform model evaluation after estimation for generative network models like ERGMs. While this should be a standard procedure regardless of the estimation method, it is an especially important reminder as the field observes the revival of non-simulation and local-simulation methods such as MPLE and CD, thanks to emerging methodological innovations and new data structures. These methods are less prone to convergence failures, which can have the hidden liability of making it harder to spot poorly-behaved models (an issue encountered in the early use of the MPLE before the availability of simulation-based evaluation, as discussed by Snijders (2002)). We recommend that researchers evaluate model adequacy by simulating networks from the fitted model and comparing their key network features and specified sufficient statistics with their observed counterparts, e.g. following the procedure of Hunter et al. (2008a), where feasible. Fortunately, simulation-based evaluation is computationally much cheaper than simulation-based estimation, as the former only requires simulation from one model while the latter needs to explore a set of models in the parameter space; thus, even when MCMLE is computationally prohibitive for evaluation, MCMC adequacy checks (a.k.a. goodness-of-fit checks) are often feasible. For sufficiently large, high-variance systems in which even this is infeasible, alternative checks are needed. Although this regime remains an open problem, conditional simulation using e.g. Held-Out Predictive Evaluation (Wang et al., 2016; Yin et al., 2019) may be one useful approach, provided that enough dependence-graph adjacent edge variables are held out simultaneously to permit detection of degeneracy. Some work has been done on bounding techniques for dichotomous networks that can in some cases rule out degeneracy without resorting to simulation (Butts, 2011); it is unclear whether similar techniques can be developed for count-valued networks, but if so such methods could prove useful where simulation is impractical. In general, evaluation for networks that are too large for complete simulation (in the count-data case or otherwise) is an important frontier for future work.

2.7 Conclusion

ERGMs, especially for valued networks, can be computationally expensive to estimate. In searching for a fast and reliable computational method, we implemented MPLE for count-valued ERGMs, and performed a comparative simulation study using three methods: CD, MCMLE, and MPLE. We found that the variance of edge values is critical in determining the performance of computational methods for Valued ERGMs, while the network size mainly influences their relative merit in computational efficiency. For small-variance networks, point estimates are easy to acquire using whichever method, while CD greatly overestimates uncertainties and MPLE underestimates them for dependence terms. All methods have similar wall-clock time. For large-variance networks, both MPLE and MCMLE offer strong performance for estimating both coefficient and uncertainties, although MPLE is an order of magnitude faster than MCMLE.

On the basis of this study, we recommend that researchers pay attention to the variance of edge values in choosing computational methods. For small-variance data, MCMLE should be the default method where feasible, although CD is useful for point estimations; MPLE is suited for large networks and high-dimensional models, especially with a large number of available processors, but caveats should be given for interpreting its standard errors for dependence terms. For large-variance networks, MPLE is a solid method, and researchers can design the size of edge sample and the structure of edge support truncation based on the computational resources at hand and the requirements of estimation quality. Our experiments also demonstrate that both CD and MPLE are useful tools for MCMLE seeding, although CD is better for simpler cases with its speed advantage and MPLE is better able to offer high-quality seeds across all scenarios.

In summary, with insights about the behaviors of each method under different network sizes and edge variances, this chapter offers a guideline for choosing and tuning computational methods for Valued ERGM estimation. The implementation of a flexible subsampled parallelizable MPLE framework is demonstrated to be a powerful tool; we envision it to empower researchers with large-variance big network data and high-dimensional model design, freeing them from the need to employ data-reduction and model-simplification compromises because of computational constraints.

2.8 Appendix: Contrastive Divergence with Different Parameters

Results in Section 2.5 showed that despite its time efficiency, CD has two limitations. First, its bias is larger than subsampled MPLE, making it a suboptimal seeding method for MCMLE (especially when the edge variance is large). Second, its calibration of uncertainty is overly conservative, making it an uninformative method for second moment estimates. This leads to the question of whether one could tweak its tuning parameters to trade its time efficiency for less biased and better calibrated estimations. In this regard, we study the quality of CD estimators when we vary CD's major parameters: steps and multiplicity. "Steps" determines the number of Metropolis-Hastings steps, and "multiplicity" determines the number of proposal for each step. The default setting in `ergm.count` package for CD is 8 steps and 1 multiplicity, which was the setting reported in Section 2.5. Here we compare that with different combinations of modified tuning parameters.

Table 2.A1 shows the bias and the wall-clock time for CD under different tuning parameters versus MPLE under its configuration for MCMLE seeding using a single processor. In this section, we only report performance for the dependence terms because of space limitation, but estimators for other covariates generally share similar patterns. For both the small-variance small network and the large-variance small networks, increasing either steps or multiplicity or both does not monotonically reduce the bias, while the wall-clock time increases monotonically as expected. For this reason, the default CD configuration seems to be the optimal choice as a seeding method for MCMLE. For the small-variance large network, although larger multiplicity does not bring less bias, increasing steps does brings monotonic decrease in bias estimation. However, CD's bias is always larger than that of MPLE, even when its wall-clock time surpasses MPLE's. This suggests that for small-variance large networks, one can increase steps to reduce the bias of CD, but it is not as efficient as using MPLE instead,

Table 2.A1: Bias and time of CD vs. MPLE

	Contrastive Divergence						MPLE
Steps	8	80	8	80	800	8000	
Multiplicity	1	1	10	10	1	1	
Small-Variance Small Network							
Bias(Mutual) (%)	0.07	0.64	1.07	1.31	0.64	0.53	0.66
Bias(Flow) (%)	1.64	1.38	0.51	0.68	2.54	3.53	0.35
Wall-clock time (s)	0.60	1.92	2.01	21.55	30.77	342.95	26.14
Small-Variance Large Network							
Bias(Mutual) (%)	0.93	0.55	1.66	0.60	0.34	0.30	0.12
Bias(Flow) (%)	2.35	1.10	1.98	1.17	0.89	0.72	0.15
Wall-clock time (s)	1.15	6.00	6.34	69.82	96.63	2177.46	769.77
Large-Variance Small Network							
Bias(Mutual) (%)	0.42	79.94	87.41	99.4	139.9	116.13	0.05
Bias(Flow) (%)	3.86	591.9	469.13	261.75	97.59	29.45	1.78
Wall-clock time (s)	5.22	12.57	13.02	94.71	81.46	876.71	18.83

Note: We use the chosen seeding setting for MPLE: 50% sample sizes for large-variance and large-network data, 75% for small-variance small network data. Wall-clock time of MPLE is from the slowest setting using one core.

which offers better estimators with less time used. To recap, increasing either or both tuning parameters for CD usually fails to yield less biased estimators, and even when that does, it is not as time-efficient as using MPLE.

Table 2.A2 displays the calibration for the two dependence terms and wall-clock time of CD under various tuning parameters versus the benchmark MCMLE, seeded by MPLE with a single processor. For both the small-variance small network and the small-variance large network, increasing multiplicity does not lead to better calibrated standard errors, but increasing steps is associated with monotonic improvement in calibration. Nonetheless, CD fails to offer as well-calibrated estimates as MCMLE in comparable time spans. This suggests that, for small-variance data, compared to increasing steps for CD, it is a better choice to directly use MCMLE for a time-efficient and well-calibrated second-moment measurement. For large-variance small-network data, although increasing steps and/or multiplicity in general alleviate its overestimation of uncertainty, increasing steps beyond a certain point can lead

Table 2.A2: Calibration and time of CD vs. MCMLE

	Contrastive Divergence						MCMLE
Steps	8	80	8	80	800	8000	
Multiplicity	1	1	10	10	1	1	
Small-Variance Small Network							
Mutual	2.11	1.04	2.73	1.75	0.22	0.05	0.04
Flow	1.88	0.93	2.57	1.51	0.19	-0.02	-0.03
Wall-clock time (s)	0.60	1.92	2.01	21.55	30.77	342.95	30.16
Small-Variance Large Network							
Mutual	2.91	1.99	3.21	2.75	0.95	0.16	-0.03
Flow	2.76	1.88	3.10	2.53	0.98	0.21	-0.09
Wall-clock time (s)	1.15	6.00	6.34	69.82	96.63	2177.46	1673.59
Large-Variance Small Network							
Mutual	2.55	0.97	1.06	0.36	-0.33	-0.56	-0.05
Flow	2.63	0.33	1.23	0.05	-0.70	-0.96	-0.11
Wall-clock time (s)	5.22	12.57	13.02	94.71	81.46	876.71	307.35

Note: MCMLE is seeded by MPLE under the specified setting in Results using one core.

to underestimation of uncertainty. Regardless, its calibration never beats MCMLE. Overall, our experiments suggest that increasing tuning parameters beyond the default setting in `statnet` does not always improve its calibration, and when it does, it is not as time-efficient as using MCMLE for calibration.

Chapter 3

Rooted America: Immobility and Segregation of the Intercounty Migration Network

3.1 Introduction

While the active drivers of migration have been extensively studied, there has been less attention to the factors that *hinder* migration – a research gap that has been called the “mobility bias” within the migration literature (Schewel, 2020). The relatively overlooked phenomenon of immobility is important in its own right, having substantial consequences for the social world. As migration influences the functioning of labor market (Hyatt et al., 2018), the landscape of stratification and social mobility (Jasso, 2011), and the sociocultural meanings in everyday lives (Bauman, 2000; Mata-Codesal, 2015), mechanisms that impede migration can have outcomes that extend far beyond the migration system itself.

Understanding immobility is an especially apt challenge in the context of the modern United

States. Long thought of as a "rootless society" (Fischer, 2002) with high geographical mobility (Long, 1991; Steinbeck, 1939), the U.S has arguably turned into a "rooted America" after a decades-long decline in migration rates (DeWaard et al., 2020; Frey, 2009). While the reality of low migration rates is clear, explanations for current population immobility are less well-developed. Macroeconomic studies have so far found that demographic and socioeconomic structures are not sufficient to explain observed levels of immobility, and neither are the business composition of labor market nor properties of the housing market (Hyatt and Spletzer, 2013; Hyatt et al., 2018; Molloy et al., 2011, 2017). A broader sociological view suggests the potential for cultural, political, and other social forces as possible explanatory factors (Tiebout, 1956; Massey and Denton, 1993; Gimpel and Hui, 2015; Stockdale and Haartsen, 2018). Moreover, the migration system has its own intrinsic feedback mechanisms that could endogenously sustain or undermine further migration (Bakewell, 2014; de Haas, 2010), which may also play a role in the population immobility. Probing the combined influence of these myriad factors requires a *systemic* treatment of the U.S. internal migration, allowing us to simultaneously examine the joint impact of social, economic, political, and demographic mechanisms on flows of migrants throughout the country. This chapter pursues such an analysis, with the objective of identifying the factors associated with both mobility and immobility in contemporary America.

Broadly, extant research on drivers of U.S. migration and immobility shares two characteristics. First, most research examines migration from an economic perspective, assuming that most, if not all, migration is *labor migration*, driven by economic incentives.[1] Yet, decisions regarding residential settlement are not purely economic (Ryo, 2013): political climate, racial composition, and urbanization of local communities are potential contributors to the phenomenon (Brown and Enos, 2021; Cramer, 2016; Massey and Tannen, 2018). This chapter incorporates the sociocultural and political perspectives into the analysis of U.S. immobility.

---

[1] As an example, Eeckhout (2004:1431) contends that "the central thesis in this chapter: population mobility is driven by economic forces."

A second dominant characteristic of the extant literature on U.S. migration is an approach that treats migration as a feature of geographical areas, examining the correlates between net migration rates into or out of states or counties and their demographic or economic characteristics. Although convenient, this practice of reducing the interconnected migration system into local features of areal units introduces two limitations. First, by aggregating across origins and destinations for migrants emigrating from or immigrating into a given area, it obscures the *interactive effects* from the sending and receiving areas, such as their political or cultural similarity and differences in employment rates. Second, it does not allow for treatment of the *internal dynamics* of the migration system (de Haas, 2010), in particular the presence of mechanisms such as return or stepwise migration, where the flow of migrants from one place to another can in turn affect the flow of migrants from that destination to others. Since migration is a relational process between places of origin and destination, and migration flows can influence each other, this chapter takes a systemic, network approach that shifts analysis from the migration rates of areal units to the migration flows *between* areal units. By leveraging migration systems theory and social network methods, we show that dissimilarities between counties are important contributors to the immobility of American society.

To advance our understanding of the social forces behind geographical immobility in modern America, we here adopt a comprehensive theoretical framework incorporating geographical, demographic, economic, political, and social influences on migration and perform a systemic analysis of internal migration as an evolving valued network of migration flows.[2] Using valued temporal exponential-family random graph models (valued TERGMs), this chapter analyzes the network of intercounty migration flows of the United States from 2011 to 2015. We identify a pattern of *segmented immobility*, where, net of other factors, less migration happens between counties with dissimilar political contexts, levels of urbanization, and racial

---

[2] By valued network (or weighted network), we refer to networks whose ties are not binary (present or absent), but are associated with a quantitative value; specifically, tie values in this study indicate the volume of migration flows between directed pairs of U.S. counties.

compositions. We probe this mechanism using an *in silico* “knockout experiment,” which suggests that in a counterfactual world without segmented immobility (but holding all other factors constant), we would expect to have seen approximately 4.6 million (27%) more intercounty migrants in the United States each year. This implies that social and political cleavages in America are substantial contributors to immobility, and potentially exacerbate growing trends towards geographical segregation. Further, we also examine the relationship between internal and international migration flows, showing that - contrary to the balkanization thesis (Frey, 1995a,b) - international migration into a county is positively associated with its overall domestic mobility, and does not promote net outflows of residents. The model also identifies the internal dynamics of migration systems (de Haas, 2010), including a suppression of what we dub “waypoint” flows (i.e., balanced in- and out-flows of an areal unit) alongside strong patterns of reciprocity and perpetuation. While the data availability constraints us to focus on understanding population immobility in the 2010s, the empirical evidence together with our proposed theoretical and methodological frameworks opens the door to unpack the long-term phenomenon of population immobility. This chapter thus joins the growing literature that grapples with the mobility bias in migration studies (Schewel, 2020), demonstrating how a comprehensive analytical framework and a systemic, network approach offers new insights about immobility, and more broadly, the dynamics of population movement among social and geographical spaces.

3.2 Theory

Existing literature defines immobility as “continuity in an individual’s place of residence over a period of time” (Schewel 2020:344). Since immobility is not only an individualistic phenomenon, but also a population and social one, here, we offer a macrosociological definition of immobility, which is a lack of population exchange between localities. Drivers of immobility, in terms of this framework, are defined as factors that *reduce* migration rates relative to what would be expected in their absence. The scarcity of migration in an immobile society has substantial impacts. Since migration is a critical channel for people to respond to fluctuations of local economy, population immobility implies a rigid labor market with lower productivity, higher unemployment rate, and more prolonged recession when experiencing economic shocks (Hyatt et al., 2018). Moreover, migration also serves as a way of improving life chances (Jasso, 2011; Weber, 1922) and coping with adverse events (Spring et al., 2021). Population immobility thus has important ramifications for social mobility, stratification, and poverty (Briggs et al., 2010; Clark, 2008; Jasso, 2011).

Immobility is not merely the flip side of mobility, but carries its own sociocultural meanings. As the aspiration-ability model argues, migration requires both aspiration to migrate and the ability to realize that aspiration (Carling and Schewel, 2018). This means that immobility is not necessarily a passive outcome of simply staying in place, but can be a conscious choice to remain. In line with this view, recent literature has begun augmenting the widely discussed notion of "cultures of migration" with the notion of "cultures of staying" that facilitate and maintain immobility (Stockdale and Haartsen, 2018). The level of population (im)mobility can in turn impact the broader social norms of a society; a mobile society may have a prevailing nomadic culture, while the dominant culture of an immobile society may be sedentary (Bauman, 2000; Mata-Codesal, 2015).

Understanding immobility is especially relevant in the American case. From the earliest observations of Tocqueville (1834) and Ravenstein (1885) to Steinbeck (1939), America has long been considered a "restless" or "rootless" society with high geographical mobility. Yet, after a decades-long decline in its migration rate, the contemporary America has arguably become a "rooted" society with considerable population immobility. However, as Herting et al. (1997:267) have noted, sociological research on U.S. mobility has "narrowed and now focused almost exclusively on mobility of a purely economic or occupational variety," with much less focus on mobility across geographic space. In migration studies, research has been historically focused on studying the social forces that lead to migration, but largely neglected the counter forces that *inhibit* people from moving, a tendency that Schewel (2020) described as the "mobility bias." A lack of research on geographical mobility in American sociology, together with the scarcity of theoretical and empirical work on immobility in migration studies, has led to gap in our knowledge regarding the mechanisms behind population immobility in contemporary American society.

3.2.1 Culture and Politics of Immobility

While the immobility of the U.S. population has received less sociological attention, economists and geographers have conducted empirical analyses on this matter (e.g., Cooke, 2013; Jia et al., 2023; Kaplan and Schulhofer-Wohl, 2017; Treyz et al., 1993). These studies have identified important connections between the labor market and migration rates, but their findings largely rely on the assumption that most, if not all, migration is *labor migration*, driven by economic incentives. The economic perspective has a fundamental role in explaining migration and immobility; the relative gains in moving, and costs associated with both transaction costs and losses of specialized local investments *are* factors that shape migration. But there also exist other factors, such as regionally specific cultural values and locally conventional ways of understanding opportunity (Carling, 2002; Carling and Schewel, 2018), as well as preferences for particular local policies or political regimes (Tiebout, 1956). Indeed, recent research on American economy has shown that over the past several decades, migration has not been effective in responding to fluctuations and shocks in labor markets (Dao et al., 2017; Jia et al., 2023). Relatedly, macroeconomic factors have not been found to have a strong correlation with migration rates in the U.S (Hyatt et al., 2018; Hyatt and Spletzer, 2013; Molloy et al., 2017). Therefore, while economic forces are important ingredients in a viable model of the migration system, a comprehensive analysis of immobility demands considerations of other social institutions.

Although thinking on internal migration in the large has been dominated by labor market considerations, sociologists have given considerable attention to other factors when studying migration at smaller scales (e.g., across neighborhoods). For instance, research on residential segregation has long identified how people with different racial identities and political beliefs become segregated from each other (Bishop and Cushing, 2009; Krysan and Crowder, 2017; Massey and Denton, 1993), including the accumulated influence of even relatively weak preferences for same-group interaction (Schelling, 1969; Sakoda, 1971); the latter can act as a powerful macro-level sorting force, even in the presence of economic or other factors (e.g., Butts, 2007). While much of this work has focused on racial segregation, more recent work has also probed segregation along political or cultural axes. For instance, Brown and Enos (2021) found that a large proportion of American adults live in neighborhoods where most residents share the same partisanship. Gimpel and Hui (2015) used a survey experiment to show that people evaluate more favorably towards properties in areas with predominantly co-partisan neighborhoods. As social cleavages might deter people from settling in places with distinct identities and beliefs, the social gaps between rural and urban areas and those among different parts of the continent such as the South and the coastal regions (Cramer, 2016; Hochschild, 2018), may also contribute to the inhibition of geographical movement. At another scale, in the contexts of international migration, migration studies have long stressed the roles of cultures and politics in shaping population mobility (Castles et al., 2013; Cohen and Sirkeci, 2011; Jennissen, 2007; Massey et al., 1999; Vögtle and Windzio, 2022; Waldinger and Fitzgerald, 2004). Following this thread, this chapter incorporates the political, racial and rural-urban structures in investigating American immobility.

3.2.2 Systemic Theories of Migration

The second characteristic of the extant literature on U.S. immobility is that studies usually view migration as a feature of geographical areas. This approach examines the characteristics of an areal unit that influence its net immigration and emigration rates, such as percentages of current residents who are immigrants and/or emigrants. It is in essence a marginal approach that sums up (i.e., marginalizes) the migration flows from/to each areal unit across all destinations/origins to describe the overall mobility of each place. The marginal approach is empirically straightforward, and has unquestionably contributed to our understanding regarding the driving forces of migration by identifying the associations between demographic and economic features of an areal unit and the scale of its population inflows or outflows (e.g., Partridge et al., 2012; Treyz et al., 1993). Yet, migration - by definition, population moving from one place to another - is *inherently* relational, having properties that cannot be reduced to the features of individual areal units. For instance, studies considering net in- or out-migration rates in isolation must choose either the sending or receiving area as focus of analysis (thereby obscuring the joint roles of areas as origins and destinations), or must merge in- and out- migration to obtain a net migration rate (which confounds inflows and outflows). Beyond the fact that every pairwise migration flow among sending and receiving areas depends on both the properties of the sender and the receiver, such studies are unable to account for relational factors, such as geographical proximity and political difference between areal units. Neither can this approach consider the interactions among migration flows, such as reciprocal population exchange ($A \rightarrow B$ & $B \rightarrow A$) arising from return migration. Probing such mechanisms requires a different theorization of the migration process, a systems theory of migration.

Such systemic thinking has a long tradition in migration studies under the umbrella of migration systems theory (MST, Bakewell (2014); Fawcett (1989); Kritz et al. (1992); Mabogunje (1970); Massey et al. (1999). A comprehensive theory that concerns the complex interactions among various elements related to migration, such as flows of people, information, (formal and informal) institutions, and strategies (Bakewell, 2014), MST identifies *interconnectivity* as a key feature of migration. As de Haas (2010:1593) summarized, a migration system is "a set of places linked by flows and counter-flows of people, goods, services and information, which tend to facilitate further exchange, including migration, between the places." The theoretical focus on flows *between* origin and destination suggests a relational analysis of migration, integrating push and pull factors in one single analytical framework (Lee, 1966). Fawcett (1989) demonstrates this with a theoretical framework of "linkages" in MST, focusing on how various linkages between origin and destination shape the migration in between. Among the linkages Fawcett (1989:677) discusses, here we focus on the relational linkages, "derived from comparison of two places." Instead of studying how a state's or a county's political climate influences its net marginal migration rate (e.g., Charyyev and Gunes, 2019; Preuhs, 1999), an analysis of relational linkages examines how the *difference* in political climates between counties influences the number of people migrating from one to the other.

Another critical implication from the interconnectivity feature of migration systems is the presence of internal dynamics of migration (Bakewell et al., 2016b; de Haas, 2010; Mabogunje, 1970). As Mabogunje (1970:16) put it, the migration system is "a circular, interdependent, progressively complex, and self-modifying system in which the effect of changes in one part can be traced through the whole of the system." Similarly, Fawcett (1989:673) argued that the migration systems framework "brings into focus the interconnectedness of the system, in which one part is sensitive to changes in other parts." This means that migration is not a pure product of exogenous social forces. It forms a system with endogenous processes, where one migration flow can promote or suppress another migration flow. For example, since migrants transmit information and social connections when they move, the migration flow from Arizona to Texas brings job information and personal contacts along, potentially inspiring migration in the opposite direction. Internal dynamics like this can lead to an endogenous accumulation of migration net of exogenous social and economic influences.

3.2.3 Migration Systems Through a Network Lens

The insight of interconnecitivity from MST resonates with that of social network analysis. Indeed, past research has employed social network analysis to study migration systems (Charyyev and Gunes, 2019; Desmarais and Cranmer, 2012b; DeWaard et al., 2012; DeWaard and Ha, 2019; DeWaard et al., 2020; Hauer, 2017; Leal, 2021; Liu et al., 2019; Nogle, 1994; Vögtle and Windzio, 2022; Windzio, 2018; Windzio et al., 2019). This school of MST, called by Bakewell (2014) the "abstract system," interrogates the macro-level migration patterns by analyzing migration networks consisting of localities (in network terms, *nodes* and migration flows between each directed pairs of localities (in network terms, *edges*).[3] Network analysis effectively captures the two critical implications of MST, relational linkages and internal dynamics of migration systems, bringing new perspectives compared to the marginal approach of migration, commonly employed in studies of U.S. immobility. Rather than viewing localities/places as units of analysis, the network approach takes migration flows between places as analytical units. This perspective preserves information regarding emigration and immigration processes, enabling analysis of how characteristics of origin and destination *interact* to influence migration flows, a relational account of linkages in migration systems. The network approach also examines the internal dynamics of migration systems, by studying the dependence structure among migration flows. The dependence structure identifies how migration flows are associated with each other, net of the exogenous contexts such as the economic and political environments. Taking the above-mentioned example of reciprocity, the network approach measures whether and to what extent, an increase of one

---

[3] Bakewell (2010, 2014) and DeWaard and Ha (2019) have debated about whether and how studies of migration networks contribute to MST. As this chapter shows, echoing Leal (2021), we agree with DeWaard and Ha (2019) that network analysis is an effective way of theorizing and testing the structures and dynamics of migration across geography; we also recognize Bakewell's critique that network analysis of migration flows is one of the many approaches to study migration systems, and that students of MST should beware the pitfall of abstract and static descriptions of migration systems. In this regard, this chapter leverages theories and empirical findings in migration studies to motivate tests about structures and patterns of migration networks. We also call for more research with different levels of analysis to triangulate our findings for a comprehensive understanding of migration and immobility.

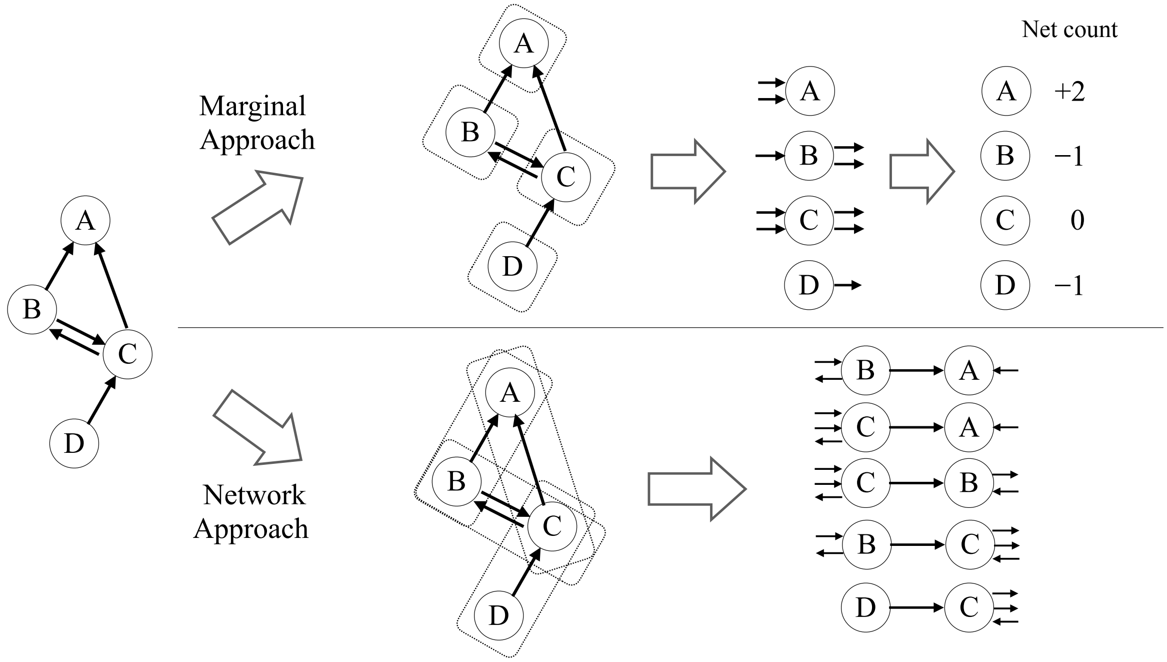


Figure 3.1: Schematic illustration of the marginal approach versus the network approach
Note: The marginal approach takes geographical areas as units of analysis, and tends to condense the in- and out- migration flows into a single number about net migration rate/count of a geographical area. The network approach takes each migration flow between a directed pair of geographical areas as an analytical unit. This approach incorporates origin and destination in understanding their joint influence on migration flows; it also preserves the local structural properties of migration flows, allowing systemic patterns to be examined.

migration flow (e.g., Los Angeles to Baltimore) is associated an increase in its opposite flow (Baltimore to Los Angeles), net of other factors. The dependence structure can further go beyond a pair of places and describes how the whole network system of migration flows are interconnected, such as how the migration inflows of Denver are associated with its outflows, which in turn serve as the inflows of another places, Dallas, Atlanta, etc. Figure 3.1 illustrates the network approach in contrast to the marginal approach.

While the network approach introduces unique perspectives overlooked by the marginal approach, its insights has not yet been fully appreciated. One notable characteristic of prior research on migration networks is the focus on the "diversity" rather than the "intensity" of migration flows (DeWaard and Ha, 2019; Leal, 2021; Vögtle and Windzio, 2022; Windzio, 2018; Windzio et al., 2019). In other words, extant research examines the *number* of migration flows rather than their *magnitudes*. This is associated with the practice of dichotomizing migration flows into two statuses, either no migrants versus at least one migrant, or few migrants versus many migrants (though Windzio (2018) and Windzio et al. (2019) divide them into more (5) statuses in some parts of their research). This approach is compatible with the common practice in social network research of approximating social relations by a binary form, facilitating the use of existing network theories and methods to describe the migration system. While analyzing the "diversity" of migration flows offers useful knowledge about the migration system, it ignores the rich information about the variation in migration magnitudes. The intensity of migration flows becomes a critical question when it comes to understanding population immobility. In particular, DeWaard et al. (2020) find that the decline of U.S. migration is not due to the decline in the diversity of migration flows (the number of county pairs with population exchange), but the decline in the intensity of migration flows (their average count of migrants). Studying the intensity of migration flows requires describing migration networks in a valued form, where the edges are not binary, but take quantitative values. Since the quantitative feature of migration intensity is critical in grappling with the question of population immobility, this chapter bridges migration systems theory and recent advances in statistical and computational methods for valued network analysis (Huang and Butts, 2024b; Krivitsky, 2012). We formally theorize the relational linkages and internal dynamics in the expressions of valued networks, developing a roadmap to quantitatively describe and test the interconnectivity of population flows.

On the side of migration systems, new theoretical insights are needed for studies of immobility. MST is not an exception from the mobility bias critique of migration theories (Schewel, 2020). As de Haas (2010) argues, MST has historically focused extensively on migration-facilitating mechanisms that lead to the perpetuation of migration flows, but largely overlooked the migration-undermining mechanisms that lead to the decline of migration flows. Building on this critique, a line of theoretical and empirical research studies why some instances of pioneer migration drive the formation of migration systems while others do not, and the endogenous mechanisms that can undermine the migration system (Bakewell et al., 2012, 2016a; de Haas, 2010). Bakewell et al. (2016a) further go beyond the MST framework, as they pursue the notion of incorporating scenarios where the migration systems fail to form or perpetuate. Unquestionably, this is a promising direction to further the theorization of migration dynamics. Yet, for our focus of internal migration in the contemporary U.S., the migration system has been in existence for generations, and is unlikely to vanish in the near future. Therefore, the migration system is still a useful research subject and perspective, where we explore the social mechanisms that immobilize population from migrating.

The network approach inspires us to consider population immobility from a relational perspective. We conceptualize the pattern of *segmented immobility*, that in a society where people cluster in geographical segments based on their cultural and political traits, immobility can occur due to people's tendency to avoid migrating towards places with divergent environments. By jointly incorporating origin and destination in an analytical framework, the relational perspective allows us to examine the influence of dissimilarity between counties on the magnitude of migrant populations moving between them, connecting population immobility with segregation and polarization. Apart from examining the pattern via a hypothesis testing lens, we further utilize the idea of "knockout experiments" broadly employed in the experimental sciences to directly quantify its contribution to immobility. Originating in biomedical research, a knockout experiment probes the functional role of a system component by removing or inactivating it, comparing normal system behavior with behavior when the component is "knocked out" (Hall et al., 2009; Vogel, 2007). In social sciences, knockout experiments are performed *in silico*, where researchers simulate the potential social outcomes when certain social forces were removed. The knockout experiment can be considered as a model-based thought experiment (*Gedankenexperiment*, (Einstein et al., 1935), in which we predict the social outcomes of interest under a counterfactual scenario where certain social effects are inoperative. In our case, we compare the total number of migrants observed in the real world to that simulated when segmented immobility mechanisms are knocked out. This theoretical exercise allows us to leverage the power of modern, generative network models to gain insights into the functioning of the migration systems.

3.3 Hypotheses

3.3.1 Relational Linkages: Political Segregation and Segmented Immobility

Decisions about migration, a behavior aiming at improving life chances (Jasso, 2011), typically come out of a comparison between place of departure and destination. Moving from current place of residence, will the destination be adaptive? One critical dimension in drawing an answer is the political environment of the origin and the putative destination communities. Rising political polarization has divided Americans along the party lines (Levendusky, 2009), where social cleavage by political ideology extends to a growing array of public opinions (Baldassarri and Gelman, 2008; DellaPosta, 2020) and choice of lifestyles (DellaPosta et al., 2015), and has lead to segregated social networks and tensions in relationships such as familial interactions (Chen and Rohla, 2018; DiPrete et al., 2011). This political alignment also happens across space, with distinct political consciousness across geographical regions, rural and urban lands, and local neighborhoods (Bishop and Cushing, 2009; Cramer, 2016; Hochschild, 2018). Recent spatial analysis on partisan isolation reveals that a large fraction of American adults lives in places where almost no one in their neighborhood votes in a manner opposed to their own (Brown and Enos, 2021). They also found that this pattern is prevalent nationwide and is a distinct pattern from segregation in other dimensions such as across racial lines. This state of affairs is also overtly recognized within American political discourse, where media outlets routinely make distinctions between "red" (conservative) and "blue" (liberal) regions, and ascribe (correctly or not) a large body of cultural and political traits to both the regions and their inhabitants (Badger et al., 2018; Wallace and Karra, 2020). To the extent that individuals are likely to both affiliate with the political culture of their area, and regard their opposites on the political spectrum with suspicion and even hostility (Iyengar et al., 2012, 2019), people may be unwilling to migrate between regions with differing political cultures. Even setting aside motivations arising from political culture, according to the public choice theory and the consumer-voter model, people should still be more willing to migrate to regions whose governments most closely match their own policy preferences (Dye, 1990; Tiebout, 1956), with those from solidly "red" areas preferring to move to other "red" areas, and likewise for those from "blue" areas. Empirical analyses using various data and methods generally confirm the existence of migration preference towards co-partisanship (Tam Cho et al., 2013; Gimpel and Hui, 2015; Liu et al., 2019), though with some counter evidence (Mummolo and Nall, 2016). Together they motivate the following hypothesis:

Hypothesis 1.1: Ceteris paribus, *the more dissimilar counties are in their average political orientation, the lower the migration flow between them.*

The limited population exchange between geographical segments with dissimilar social environments, or what we call *segmented immobility,* may not be unique to the political dimension, but would rather be a pervasive pattern arising from people's evaluation of places along multiple dimensions. One of the underlying mechanisms that can lead to such a pattern is homophily. Homophily refers to people's tendency to be connected to and interact with those similar to themselves in various characters such as racial and ethnic identity, religious belief, political ideology, personality, and normative inclination like altruism (DiPrete et al., 2011; Leszczensky and Pink, 2019; McPherson et al., 2001; Moody, 2001; Smith et al., 2014; Wilson et al., 2009). Homophily occurs not only within personal networks, but is also a spatial phenomenon, where people tend to live close to others with similar racial identity, economic background, or political ideas (Bishop and Cushing, 2009; Massey and Denton, 1993; Intrator et al., 2016). A social process that can give rise to this spatial pattern is that residents choose to migrate towards places where people similar to them concentrate, but avoid destinations with identities different from their own (Crowder et al., 2012; Massey et al., 1994b; Schelling, 1969). Although literature about this residential sorting process focuses primarily on mobility among neighborhoods in urban areas, we argue that a similar process may also work at a larger scale. When choosing a county to reside in, people may favor places with a significant presence of their co-ethnics and those that host like-minded residents. Likewise, opportunities to migrate may be turned down if they would lead to settings in which the mover would find themselves socially isolated or targets of discrimination.

Segmented immobility can also arise in more subtle ways: even if individuals do not avoid living with dissimilar others, they may exclude potential migration destinations that are not able to offer the lifestyle and cultural consumption they are used to. Moving from Manhattan to rural Texas, the New Yorker would miss the coffee shop at the street corner, while a Texan migrating in reverse might feel nostalgia for the country music scene back home. Hence, migration between rural and urban areas, and across culturally different states is likely to be disfavored. Racial demographics can also be a determinant of the cultural and economic conditions of a place, where a racially diversified area not only offers a diversity of cultural affordances (as reflected by cuisines and music genres, for example), but also provides vital economic opportunities and ethnic capital for ethnic minorities (Fernández-Kelly, 2008; Lee and Zhou, 2017; Zhou, 1992). Similarly, migrants from rural counties might find themselves excluded from jobs in urban areas because they demand skills hard to obtain in their rural hometown, potentially leading to circulation of poor rural migrants among non-metropolitan counties (Lichter et al., 2022). These together suggest an economic dimension to segmented immobility, in which migration between dissimilar places is suppressed when these places have different economic structures, making it difficult for migrants to utilize human capital accumulated in their place of origin. As services, cultural activities, and modes of production become specialized to a local social ecology, those adapted to both producing and consuming within that ecology will find it increasingly difficult to utilize opportunities in ecologically distinct localities. Together, these mechanisms lead to the following hypotheses:

Hypothesis 1.2: Ceteris paribus, *the more dissimilar counties are in their levels of urbanization, the lower the migration flow between them.*

Hypothesis 1.3: Ceteris paribus, *the more dissimilar counties are in their racial compositions, the lower the migration flow between them.*

The hypothesis of segmented immobility is based on the assumption that most residents and migrants identify themselves with their current residence, which is also the place of departure. However, if we were to suppose that the majority of the migrating population moved to *escape* their current residence in favor of one more to their liking, then migration flows would preferentially occur between dissimilar areas; this would lead to "mobility across segments," in contrast to "segmented immobility." This type of process was proposed by Tiebout (1956) as a mechanism of political sorting, and at the micro-level similar processes have been occur in personnel turnover (Krackhardt and Porter, 1986) and cascade-like relocation phenomena (Schelling, 1978). We contend that such sorting flows are unlikely to be the major force of the contemporary internal migration in the United States. This is because research has not documented substantive social changes that drove massive redistribution of American population since the fading of the Great Migration of Black Americans in 1970s (Sharkey, 2015; Tolnay, 2003), and the continuing decline of internal migration for the past decades seems to suggest a scenario of equilibrium, or "an inflection point" (Molloy et al., 2011): 173). Analyses of voting behaviors also reveal that internal migrants tend to hold political orientations consistent with those of their origins (Preuhs, 2020). Nevertheless, we consider it as a competing hypothesis to the segmented immobility hypotheses above, and will directly test them in our analysis.

3.3.2 Internal Dynamics: Reciprocity and Perpetuation

The network approach also brings the opportunity to formally examine the interrelationships among migration flows themselves, thereby revealing the internal dynamics of the migration system. This is particularly true for the valued network models used here, which allow us to examine quantitative questions that go beyond the simple presence or absence of migration. Here, we focus on several mechanisms motivated by prior theory on migration behavior at the micro-level, which lead to hypotheses regarding interdependence among macroscopic migration flows.

We begin by considering the relationship between one migration flow (e.g., from Seattle to Austin) and its opposite flow (e.g., from Austin to Seattle). As has been argued by the transnationalism school in the context of international migration, migration is not a one-way process, but an enduring reciprocal exchange of people, goods, and cultures between sending and receiving countries (Schiller et al., 1995; Waldinger, 2013). These same mechanisms could also apply to movement within countries: in his classic work, Ravenstein (1885:187) documented the "universal existence" of "counter-currents of migration" between counties in the United Kingdom, where population not only moved from agricultural areas to commercial and industrial areas, but each of these migration currents corresponded to a current running in the reverse direction. Considering that migration control policies suppress the circulation of international migrants between states (Czaika and de Haas, 2017; Massey et al., 2016), we expect even stronger reciprocity of migration flows in the context of internal migration in the U.S., where there is no state control over migration. The reciprocity can arise from the sharing exogenous properties of the bidirectional flow; for example, geographical proximity is a driver of reciprocal population exchange, as it facilitates migration in both directions. Nevertheless, we argue that reciprocity is also an internal dynamic of the migration flow system, such that net of exogenous factors, a larger migration flow in one direction is still associated with a larger migration flow in the opposite direction.

The endogenous, systemic pattern of reciprocity could result from at least two micro-mechanisms in the American migration system. First, migration in one direction actively motivates the flow in the opposite direction. Migrants bring information and social connections from their origin to destination, inspiring and facilitating migration in the opposite direction. Second, return migrants participate in flows in both directions, contributing to the positive association between the pair of flows. For example, Spring et al. (2021) find family ties to be a decisive factor for people separated from their spouses or cohabiting partners to return to their hometowns. von Reichert et al. (2014a,b) show that migrants returning from urban to rural areas are mainly driven by social connections rather than economic opportunities, and they usually bring people in their family network along when they return. Given the plausibility of both mechanisms, we posit the following macro-level hypothesis:

Hypothesis 2.1: Ceteris paribus, *the flow of migration from county A to county B increases with the flow of migration from county B to county A.*

As is implied above, an important feature underlying the macro-level pattern of reciprocity is the presence of (interpersonal) migrant networks that link persons in the sending and receiving regions, so as literature on transnationalism points out (Lubbers et al., 2020; Mouw et al., 2014; Verdery et al., 2018). Migrant networks, according to the definition of Massey et al. (1993:448), "are sets of interpersonal ties that connect migrants, former migrants, and nonmigrants in origin and destination areas through ties of kinship, friendship, and shared community origin." We have argued that, theoretically, migrant networks should contribute to the reciprocity of migration-flow networks, by migrants bringing resources to destination and triggering population moving in the opposite direction, and by motivating return migrants moving between regions in both directions. Yet, reciprocity is not the only pattern that emerges from migrant networks. As the cumulative causation theory argues, the formation and development of migrant networks are a key contributor to the perpetuation of migration flows, which suggests the presence of inertia (aka. a positive association) of the same migration flow over time (Massey, 1990; Massey et al., 1993). Specifically, migrants not only bring information and social connections of origin to their destination, triggering migration in reverse, but also take those kinds of resources from destination back to their origin, by returning home or via communication with nonmigrants back home; this lowers the costs and potentially raises the aspiration of migrating to the same destination, making future migration more likely to happen (Garip, 2008; Garip and Asad, 2016; Liang et al., 2008; Liang and Chunyu, 2013; Lu et al., 2013; Massey et al., 1994a; Palloni et al., 2001). Therefore, we hypothesize the perpetuation of migration flows in the system:

Hypothesis 2.2: Ceteris paribus, *the flow of migration from county A to B increases with the past flow of migration from county A to county B.*

3.3.3 Waypoint Flows

We now turn to the internal dynamics at the level of triads, i.e., among three localities (Davis and Leinhardt, 1972). Specifically, we examine the *waypoint structure* in the migration flow networks. Similar to a layover airport that mainly serves connecting flights, the "waypoint" is a place where its scales of migrant inflows and outflows are similar to each other. Demonstrated in Figure 3.2, County A, B, C have the same amount of associated migration events in total (six), but their distributions of immigration and emigration are different. County A is an example of waypoint, where inflows and outflows are evenly distributed, while County C is a counter-example that has few inflows but many outflows, and County B is in between. The difference can be represented by the measure of *waypoint flow*, which is the total amount of migration flows moving in and out of a focal place. When we hold constant the total number of migration events, a high volume of waypoint flow represents a high level of equality between their inflows and outflows. In Figure 3.2, the volume of waypoint flows for County A, B, C are three, two, one, respectively, indicating that County A has the most

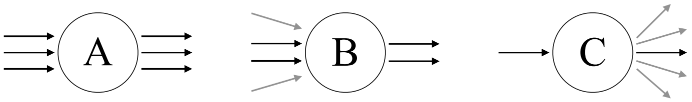


Figure 3.2: Waypoint flows
Note: County A, B and C have the same number (six) of associated migration events, but their levels of equality in the in- and out-migration flows vary. This is reflected on their volumes of waypoint flow, three for the most equal County A, two for the medium equal County B, and one for the least equal County C.

balanced inflows and outflows, followed by B and C.

Waypoint flows can arise from chain-like migration processes (Leal, 2021), such as stepwise migration and relay migration. Stepwise migration refers to movements of migrants that pass through at least one waypoint before reaching the final destination (Conway, 1980). Originally theorized in the classic piece of Ravenstein (1885), stepwise migration has been widely documented to happen under various social contexts (Freier and Holloway, 2019; Paul, 2011, 2017; Riddell and Harvey, 1972), including internal migration in the United States (DeWaard et al., 2016). Stepwise migration usually happens when the final destination is not directly reachable because of the high financial burden or the hardship in acquiring visas for international migration; migrants respond to this challenge by first migrating to waypoints that facilitate their accumulattion of capital of various kinds before moving to their ultimate stop (Paul, 2011). Another migration process that gives rise to waypoints is relay migration, where exodus of local residents leave vacancies in the labor market that attract inflows of migrants (Durand and Massey, 2010). Relay migration can also happen in the reverse order, where the influx of migrants triggers outflows of local residents (Leal, 2021). The key difference between stepwise migration and relay migration is that the former is about the same migrant taking a multiple-step move, but the latter involves different populations participating in the inflows and outflows of waypoints.[4] The two processes are not distinguishable in aggregate migration flows, but both reflect the interconnectedness of

---

[4] We thank an anonymous reviewer from the *American Sociological Review* for pointing out this distinction.

the migration system, where the change of one migration flow could alter another via their shared connection at the waypoint.

While existing literature has studied the migration processes that can generate waypoint flows, less is known about their prevalence in the migration systems. This knowledge gap drives us to further theorize chain-like migration processes by considering them against other migration processes. Since migration is an arduous undertaking with substantial risks, costs and barriers (Carling and Schewel, 2018; Liang et al., 2008; Schewel, 2020), prolonging one-step migration into stepwise is not a desirable choice unless necessary. Compared to international migration, internal migration in the U.S. is usually more affordable and less constrained by state regulations; an American internal migrant is thus less likely to opt for stepwise migration than a Filipino who wishes to settle in Spain. Relay migration is not a universal pattern, either. It requires substantial inflows or outflows that can alter the local labor and housing market or socio-political contexts to trigger further migration flows. This means that waypoint flows arising from relay migration is conditioned on uncommon incidents such as major economic shocks or environmental disasters that bring mass population movements.

Moreover, a *deficit* in waypoint flows can also be a structural signature of inequality in migration flow networks, where the majority of counties either receive many migrants but send few, or send many migrants but receive few. This imbalance between in- and out-migration flows can arise when the difference in the level of attractiveness across places remain unaccounted for; in this case, a county is either popular so to attract and retain migrants, or the reverse. A lack of waypoint flows can also occur endogenously. For instance, potential migrants may take current levels of migration rate as social or economic signals about the long-term desirability of an area, and adjust their own decisions accordingly. This tendency creates a feedback loop in which influx of migrants to an area leads potential out-migrants from the area to instead remain, which in turn feeds an imbalance between in- and out-migration ($in > out$) that motivates yet more potential migrants to move in. By turns, this same mechanism may lead to a Schelling-like exit cascade (Schelling, 1978), in which an initial out-migration shock both encourages further exit from those now in the location and makes the location appear less-desirable to potential in-migrants, thus leading to poorer in/out balance ($in < out$), and further net out-migration.

Clearly, then, there are interesting and plausible hypotheses in both directions. For simplicity, we hypothesize a high-waypoint scenario, reflected by a balanced distribution of inflows and outflows:

Hypothesis 3: Ceteris paribus*, the inflows of migration to a county increase with its outflows.*

It should be noted that the waypoint flow is a network structure related to but distinct from the transitive hierarchy studied in some international migration network research (Leal, 2021; Windzio, 2018). Both are triadic structures concerning migration flow among three places (i, j, k). The waypoint flow is the backbone of the transitive hierarchy, as the former considers migration flows of $i \rightarrow j$ and $j \rightarrow k$, while the latter involves the co-presence of $i \rightarrow k$ flow. This means that networks with a lack of waypoint flow will have few closed transitive triads $(i \rightarrow j, j \rightarrow k, i \rightarrow k)$.[5] We thus focus on the more fundamental waypoint flow structure to explore the more basic form of endogenous mechanism in the migration network.

3.3.4 Internal Migratory Response to Immigration

Lastly, this chapter considers the relationship between international migrant (i.e., immigrant) inflows and internal migrant flows in the United States. Debates about the impact

---

[5] This is because transitive hierarchy is a network structure built on waypoint flow, and an underrepresentation of the former necessarily implies an underrepresentation of the latter. It is possible that in this circumstance there can be net tendency for waypoint flows to be transitively rather than cyclically closed *where they occur*. But one will still see fewer transitive closures (as there are fewer paths to close in the first place) than one would expect by chance. Put another way, standard transitivity effects measure the overrepresentation of both waypoint flow and transitive closure, not merely the latter.

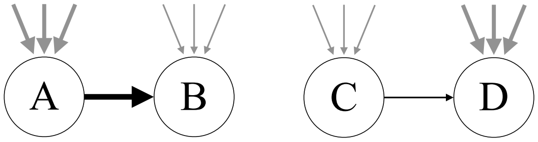


Figure 3.3: Hypothesized relation between internal and international migration
Note: vertical grey arrows denote international immigration flows and horizontal dark arrows denote internal migration flows. Arrow width denotes the magnitude of migration flows. According to the hypothesis of Frey (1995a), larger population are expected to migrate from County A that has high immigrant inflows towards County B that has low immigrant inflows, while less population would leave County C that has low immigrant inflows towards County D that has high immigrant inflows, net of other factors.

of immigration on internal migration provoked much research in 1990s, which provided insights about the demographic and economic influence of immigration, the structure of labor markets, and the social cohesion of American society. Frey (1995a) hypothesized that immigration to the U.S. would lead to demographic balkanization, in which immigrant inflows trigger outflows of internal migrants and deter their inflows. Figure 3 visualizes this hypothesis from the perspective of internal migration flows, where larger population are expected to migrate from County A that has high immigrant inflows towards County B that has low immigrant inflows, while less population would leave County C that has low immigrant inflows towards County D that has high immigrant inflows, net of other factors. This mechanism was proposed to lead to a "balkanized" regionalization of the U.S., with immigrants and natives increasingly segregated in different regions.[6] Empirical findings were inconclusive about the relationship between internal and international migration flows, with some supporting evidence for Frey's (1995a) hypothesis (Borjas, 2006; Frey, 1995a,b; White and Liang, 1998), and other opposing evidence (Card, 2001; Kritz and Gurak, 2001; Wright et al., 1997). This chapter revisits this debate with new data about migration in 2010s of all U.S. counties. Following Frey's (1995a) proposal, we hypothesize, from the perspective of internal migration flows, that:

---

[6] Since the phrasing of "balkanization" can be construed to carry certain normative connotations regarding immigration, we follow the practice in Kritz and Gurak (2001), and phrase the phenomenon as the internal migratory response to immigration.

Hypothesis 4: Ceteris paribus, *an internal migration flow increases with international immigration inflow in the sending county, but decreases with international immigration inflow in the receiving county.*

3.4 Data and Methods

3.4.1 Valued TERGMs

We use the valued temporal exponential-family random graph models (valued TERGMs) to study the intercounty migration-flow network within the United States. Exponential-family random graph models (ERGM) offer a flexible framework that describes the probability of observing certain network structure as a function of their nodes' covariates, edges' covariates, and the dependence structure among edges (Hunter et al., 2008b; Wasserman and Pattison, 1996). This empowers us to simultaneously model the characteristics of areal units (nodes' covariates), the relational linkages (edges' covariates), and the internal dynamics (dependence structure) hypothesized to characterize migration-flow networks. Previous research has employed ERGMs in a wide range of social network settings, including friendship networks in schools (Goodreau et al., 2009; McFarland et al., 2014; McMillan, 2019), inmate power relationships in prison (Kreager et al., 2017), collaboration networks in firms (Srivastava and Banaji, 2011), online social networks (Lewis, 2013, 2016; Wimmer and Lewis, 2010), and various types of gang networks (Lewis and Papachristos, 2019; Papachristos et al., 2013; Smith and Papachristos, 2016). While most studies model social relations as binary networks (i.e., encoding only whether or not relationships exist), it is more accurate and informative to model migration-flow systems as valued networks, where edges represent the size of population migrating between county pairs. Although valued ERGMs (VERGMs) are to date less well-studied than binary ERGMs, we employ the count-data ERGM framework of Krivitsky (2012) to capture migration rates in a quantitative fashion. Our model also incorporates temporal effects (the perpetuation pattern), making it a valued temporal ERGM, or valued TERGM).

We detail the model setup, computation methods and procedures in Appendix (3.7.2). We also develop and report a model adequacy check for VTERGMs, detailed in Appendix (3.7.5).

3.4.2 Knockout Experiments

Exploiting our ability to quantitatively model the magnitude of migration flows using VTERGMs, we perform *in silico* "knockout experiments" to show the impact of modelled social mechanisms in influencing the size of the migrant population, tackling the question of how particular social forces give rise to immobility. Originating and widely used in the experimental sciences (Hall et al., 2009; Vogel, 2007), this way of thinking has also been applied in the social sciences (e.g., Han et al., 2021; Lakon et al., 2015; Xie and Zhang, 2019), especially in the context of agent-based modelling (Miller and Page, 2009). For social science research, the knockout experiment can be considered as a model-based thought experiment (*Gedankenexperiment*, Einstein et al., 1935), where we use models to predict social outcomes of interest (e.g., total number of migrants) under a counterfactual scenario where certain social mechanisms are removed (e.g., the political segmentation effect) while other factors are held constant. A flexible and powerful tool, *in silico* experiments require careful research design before its utilization, as elaborated in Appendix (3.7.3). This approach is particularly powerful for nonlinear, systemic models like those used here, where seemingly small, local effects can have global consequences.

Our knockout experiments are performed as follows. Starting with a VTERGM calibrated using empirical migration data, we compute the total expected number of intercounty migrants when either the political segmentation mechanism (*per se*) or all of the three segmentation mechanisms (jointly) are knocked out, and compare this number with the observed migrant population size. The differences in total migrant population between these scenarios thus offer an insight about the scale of mobility suppression from these segmentation mechanisms - i.e., if we could "turn them off," what would we hypothetically expect to see? The counterfactual scenario was simulated by the Markov chain Monte Carlo (MCMC) algorithm based on the network model with zero coefficient values for the specified knockout social effects, or equivalently setting all dissimilarity measures equal to zero.[7] The simulated migration system can be interpreted as a representation of a counterfactual scenario where all U.S. counties had the same political climate, urbanization level, and racial composition, generating zero dissimilarity among all county pairs. It can also be interpreted as a representation of a counterfactual scenario where segmentation had no influence on migration (under certain model assumptions), as indicated by zero coefficient values for dissimilarity effects.[8] Since the network model specifies the dependence structure between migration flows, it accounts for both direct impacts of the segmentation between each county pair on their own migration flows, and the indirect impact arising from the internal dynamics of migration systems that spillover this exogenous impact. It thus offers a systemic depiction of the segmented immobility pattern.

3.4.3 Data

We analyze the intercounty migration flow data for from the American Community Survey (ACS) (U.S. Census Bureau, 2019). As a political unit with reliable demographic and economic data, counties serve as a level of geographical area that effectively describes the

---

[7] We also simulated networks using the full model (without knockouts), and calculated the difference in the total migrant size between the full-model simulation and the observed, as a measurement of bias introduced in the procedure. We then corrected the total population sizes in knockout scenarios by extracting that difference. As the difference is 0.7% of the observed migration volume, corrected and uncorrected estimates are nearly identical.

[8] We offer further discussions about designing and interpreting the knockout and other *in silico* experiments in Appendix (3.7.3).

social contexts of residents such as political environments and rurality (Lobao and Kelly, 2019; Mueller and Gasteyer, 2023; Schroeder and Pacas, 2021). Movement across a county boundary is a frequently-used definition of internal migration in the literature (Brown and Bean, 2016; DeWaard et al., 2020; Hauer, 2017; Partridge et al., 2012). Administered by U.S. Census Bureau, ACS surveys respondents' location of residence one year ago and estimates the population size that migrated between each pair of counties each year.[9] Their released data reports the averaged annual migrant counts in a five-year time window in order to have enough monthly samples for reliable estimation at the inter-country level. The outcome of interest is the count of migrant population flowing between 3,142 counties in the United States during 2011-2015.

The explanatory variables are from 2010 United States Census and ACS 2006-2010 (U.S. Census Bureau, 2011a,b). Specifically, the intercounty distance was calculated based on the 2010 Census by National Bureau of Economic Research (2016). We use presidential election turnout in 2008 to indicate the political climate of each county (MIT Election Data and Science Lab, 2018). Data sources for each covariate are listed in Appendix (3.7.1).

3.4.4 Variables

Dependent edge variable. The model predicts the count of migrants moving between each directed pair of counties during 2011-2015 from the American Community Survey. Because the count-valued ERGM effectively operates through a logarithmic link (see Krivitsky (2012), we are able to directly predict untransformed migrant counts in the model.

---

[9] Another dataset that reports counts of county-to-county migration flows is offered by the Internal Revenue Service (IRS) (Hauer and Byars, 2019). While ACS is a nationally representative demographic survey, the representativeness is a potential concern of the IRS data, as it only contains people filing tax returns, and therefore are not representative of the elder, the low-income, and the immigrant populations. Further, the IRS data of the post 2011-2012 period currently suffers from systemic problems that are not yet resolved (DeWaard et al., 2022). Nonetheless, the IRS reports migration data annually, and can be useful for fine-grained dynamic analysis of migration before 2011.

Dissimilarity score for segmented immobility. The segmented immobility thesis contends that less migration happens between places with different political climates, levels of urbanization, and racial compositions. To test the hypotheses, we measure the dissimilarity within each pair of counties along these dimensions as edge covariates for migration flows. We use the L1 distance measure, or what was called the dissimilarity score in social segregation literature (Massey and Denton, 1988). For difference in political climates, we follow Liu et al. (2019) and calculate the absolute difference in percentage of votes for the Democratic candidate in the 2008 presidential election, a behavioral measure of partisanship.[10] For levels of urbanization, we calculate the absolute difference in percentage of population residing in rural areas, a standard urbanization measurement reported in 2010 Census. For racial/ethnic composition, we use a function of the sum of absolute differences in population share for each racial category. Formally, we describe relationship between counties A and B by

$$R_{AB} = \frac{1}{2}\sum_{i=1}^{n}\left|\frac{P(A)_i}{P(A)} - \frac{P(B)_i}{P(B)}\right|$$

where R_{AB} is the dissimilarity score of racial composition between county A and county B, $P(A)$ is the total population size of county A and $P(A)_i$ is the population size of the i-th racial group in county A. We follow the Census to consider the following five racial/ethnic categories, Hispanic or Latino, Non-Hispanic Black or African American, Non-Hispanic Asian, Non-Hispanic White, and population with the other racial identifications. The difference is divided by two to make the theoretical value of the score range from 0 to 1. The higher the dissimilarity score, the more different the two counties are in the measured dimension, and

---

[10] Given how Hawaii and Alaska calculate their election results, we conduct the following operations to map their local election data to counties. Since Kalawao County, HI is regarded as a part of Maui County, HI for election purposes, we input the election results of both counties with their pooled results. Election results in Alaska were reported by election districts rather than counties. We used the map to match election results of the 40 districts with the 28 counties. The result of a county was input with that of its district if the county was affiliated with one single district. We take the mean of the results of the districts that a county spans if the county is affiliated with multiple districts. The approximation would underestimate the political difference between counties, but the bias should be minor as the affected county takes less than 1% of the sample. We thank the election offices of Hawaii and Alaska for clarification and maps of the election districts during 2002-2013 in Alaska.

the less migration is expected according to the hypotheses.

Network covariates. We utilize the mutuality statistic in the ergm.count R package to measure reciprocity in migration flows (Krivitsky and Butts, 2013).The reciprocity statistic calculates the summation of minimum value of each pair of edges by dyad. Formally, $g_m(y) = \sum_{(i,j)\in\mathbb{Y}} min(y_{ij}, y_{ji})$, where $\mathbb{Y}$ denotes the set of all i, j pairs. A positive coefficient for indicates reciprocity within the network, such that a large migration flow is more likely to have a larger counter current rather than a smaller one, *ceteris paribus*.

The model also includes the number of migrants in the past 5-year window during 2006-2010 in log scale from ACS as an edgewise covariate, to account for the association of migration flows over time, utilizing the temporal feature of TERGMs. A positive coefficient for this term suggests the perpetuation of migration flows over time, while a negative coefficient suggests negative dependence between past and present flows.

Waypoint flow is captured by the summation of the volumetric flow for each county in the network. Intuitively similar to the notion of the flow volume “through” or “across” an areal unit in the field of fluid mechanics, the flow associated with a given unit is the minimum of its total inflows and its total outflows.Formally: $g_f = \sum_{i\in\mathbb{V}} min\{\sum_{j\in\mathbb{V},j\neq i} y_{ij}, \sum_{k\in\mathbb{V},k\neq i} y_{ki}\}$, where $\mathbb{V}$ is the set of all vertices/nodes (counties), and y_{ij}, y_{ki} are values of the edge from county i to j and k to i, respectively. The term is similar to the 2-paths or mixed-2-stars in binary ERGMs, which is the number of times a node receives an edge and sends another (Morris et al., 2008). A positive coefficient for the flow term indicates that the observed network has larger volumes of waypoint flows than would be expected given all other mechanisms and covariates specified in the model, suggesting a relatively equal distribution of in- and out-migration flows across counties, and a negative coefficient would indicate otherwise.

To examine the relationship between internal and international migration flows, for each intercounty migration flow, the model measures its associations with the total immigrant inflows of its sending and receiving counties in the same time window (2011-2015). The international immigrant population is transformed by taking the natural logarithm.

Demographic covariates. The model also accounts for areal characteristics that might influence intercounty migration. These include demographic characteristics of the sending and receiving counties, from basic geo-demographic statistics to demographic compositions.

Classic models from spatial econometrics (a.k.a. the gravity model) suggest that migration rates are positively associated with the population sizes of the sending and receiving regions, but negatively associated with their distance, with a general power law form (Boyle et al., 2014; Poot et al., 2016; Zipf, 1946, 1949). Such models can be expressed by a linear combination of population and distance in the log space. Formally,

$$log(M_{AB}) = \beta_0 + \beta_1 log(P_A) + \beta_2 log(P_B) + \beta_3 log(D_{AB}) + \varepsilon$$

where M_{AB} is the migration volume from A to B, P is the regional population, D is the interregional distance, β is a covariate vector, and ε is the residual. Almquist and Butts (2015) suggest that this may arise from the volume of interpersonal contacts between regions, which also frequently scales in power law form. Although we do not use a regression model of this type here, we emulate this class of effects within our own model by incorporating (1) the log populations for the sending and receiving counties and (2) the log distance between counties (in kilometers) as predictors of intercounty migration rates; this means that our models can be considered as an extension of the gravity model. We also include population densities of sending and receiving counties (in thousand people per squared-kilometer), since Cohen et al. (2008) has shown that population density is a critical factor in predicting international migration flows. We use data from the 2010 Census for the covariates listed above.

For demographic composition, the model first considers the age structure of sending and receiving counties, as Kim and Cohen (2010) found that migrants are more likely to leave younger countries towards older countries in the context of international migration. Using the 2010 Census, the potential support ratio (PSR) equals to the ratio of population aging 15-64 over population aging 65+, which is the inverse of dependency ratio in demography literature; the higher PSR, the younger the population.

Racial composition could influence the mobility of population as well, as extant literature found different patterns of internal migration between racial groups (Crowder et al., 2012; Sharkey, 2015). Hence, besides the dissimilarity of racial composition between counties, we also consider the racial composition of the sending county to account for the varying mobility of different groups, as measured by the proportion of each racial category in the population.

Economic covariates. Economic structures of origins and destinations could potentially influence their migration flows. Since renters on average are more mobile than house owners (Frey, 2009; Molloy et al., 2011) even after controlling for demographic and socioeconomic factors (Jia et al., 2023), the model includes the percentages of housing units occupied by renters for both origin and destination, using 2010 Census data. The model also controls the percentage of population with a college degree using the 2006-2010 ACS. This is because human capital may offer greater ability and opportunities for migration, and previous analysis found that population with higher education attainments have higher migration rates in the U.S. (Frey, 2009).

Neoclassical economic theory predicts that people migrate towards economic opportunities (Massey et al., 1993; Todaro, 1976). The theory also predicts that regions with more economic opportunities will send more migrants, since their population have more capital to finance their migration (Massey and Espinosa, 1997). We thus include the unemployment rate of the origin, and the difference in the unemployment rate between the destination and the origin. In combination of neoclassical economic theory and the aspiration-ability model

(Carling, 2002), we hypothesize that more migration will come from counties with lower unemployment rate given their greater ability to move, and more migration will happen when the destination has lower unemployment rate than the origin, offering more economic opportunities and higher aspiration for migration. Similarly, the models incorporate the logarithm of median monthly housing costs of the origin and the difference in log housing costs between destination and origin.

Geographical covariates. Besides distance between counties, the model also controls for regional differences in mobility. Previous research found that migration rates and their trends in different parts of America vary significantly (Frey, 2009). We believe that the regional difference may not be fully explained by difference in social contexts indicated by the covariates above. Dummy variables are created to indicate whether the origin and destination is in the West, the Midwest, the South, with the Northeast as the reference group, based on the definition of U.S. Census Bureau (2013).

Administrative boundaries are likely to influence migration flows as well. Charyyev and Gunes (2019) found that, marginally speaking, the majority of intercounty migration in the U.S. happens within a state, and in this chapter we further examine whether state boundary influences migration flows after controlling for distance and dissimilarity between counties. Intrastate intercounty migration could be more prominent than cross-state migration because compared to intrastate migration, the cross-state migration creates extra burdens ranging from adaptation to unfamiliar legal and cultural environments, to navigation of administrative procedures such as change in occupational licensing for workers in certain occupations (Johnson and Kleiner, 2020). Yet, the opposite hypothesis is plausible under the consumer-voter model, which contends that people vote by their feet (Dye, 1990; Tiebout, 1956); as means of pursuing favorable policies, cross-state migration is more effective if people migrate to seek lower tax rates or more welcoming policies and climates for immigrants (Preuhs, 1999; Schildkraut et al., 2019). The model creates a dummy variable indicating whether the two counties are affiliated with the same state. A positive coefficient would suggests that intrastate intercounty migration is more prominent, and a negative coefficient suggests that inter-state migration is more prominent.

Variable Setup. We report two models in the following results section. The first model contains every covariate except the rural dissimilarity score, which is later included in the second model, the full model. Since the level of urbanization is strongly associated with their political environment, comparison between the two models could reveal how much of the total effect of political dissimilarity might be explained by their difference in the level of urbanization. Besides the sum term serving as an intercept, we add to models a term that counts the number of nonzero dyads of the network to account for the zero-inflation of migration flow data (Krivitsky and Butts, 2013). Its negative coefficients in Table 3.1 indicate the sparsity of migration flow network, that a county pair is more likely to have no migrants moving between than otherwise, even after controlling for all the covariates in the model. Summaries of descriptive statistics and data sources are attached in Appendix (3.7.1).

3.5 Results

3.5.1 Bivariate Analyses of Migration and Political Division

To explore the pattern of segmented immobility by political orientation, we first perform bivariate analyses between intercounty migration and political division, as visualized in Figure 4. We divide counties into two broad groups, Democratic counties and Republican counties. Democratic counties are counties where the Democratic candidate (Obama) received more votes than the Republican candidate (McCain) in the 2008 presidential election, and vice versa for the Republican counties. The sociogram in Panel A of Figure 4 visualizes the

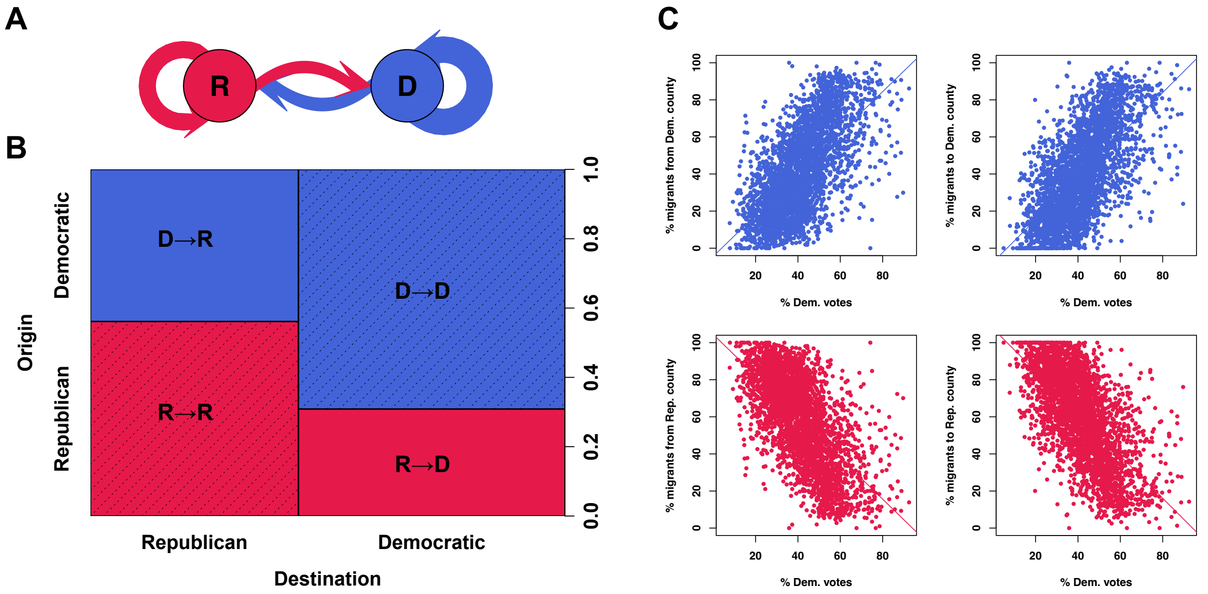


Figure 3.4: Immobility from political division

Note: The sociogram (A) represents the magnitude of migration flow within and between Democratic counties (node D in blue) and Republican counties (node R in red), which is proportional to the width of the edge. The spineplot (B) represents the magnitude of migration flow within and between the two groups by the area of each block. The shaded blocks represent migration within each group. Scatterplots (C) show the relationship between percentage of Democratic votes in 2008 of a county and the composition of its in-migrants and out-migrants. The lines are fitted bivariate linear regression lines.

magnitude of migration within and between Democratic and Republican counties, which is proportional to the width of edges. The sociogram shows that migration flows within each group has thicker edges than flows between, suggesting that more migration happens from one Democratic county to another, or from one Republican county to another, than between a Democratic county and a Republican county. The spineplot in Panel B represents the magnitude of migration flow within and between groups by the area of each block. The shaded blocks are migration happening within Democratic or Republican county groups, suggesting again that more migration happens on either side of the party line than across it. The color of each block indicates whether the origin of the migration flow is from a Democratic (blue) county or a Republican (red) county. The spineplot indicates that only 31% of the migrants moving into a Democratic county come from a Republican county, and just 44% of the migrants moving into a Republican county come from a Democratic county.

Panel C of Figure 4 visualizes the relationship between the percentage of the Democratic votes in the 2008 election and the composition of the in-migrants and out-migrants for each county. The upper left panel shows that the higher the Democratic vote in 2008, the larger the proportion of migrants coming from a Democratic county, and the smaller the proportion of migrants coming from a Republican county, as shown in the lower-left panel. Similarly, the right-hand column suggests that a larger share of 2008 Democratic votes within a county is associated with a larger proportion of out-migrants moving to a Democratic county, and a smaller proportion to a Republican county. Overall, the figures reveal a clear and strong pattern of political sorting, where less population migrate between counties with distinct political environments than those with similar political environments.

3.5.2 Segmented Immobility

The bivariate analysis is suggestive that intercounty migration is immobilized by political divisions in the United States. We further examine this using VTERGMs that incorporate the demographic, economic, geographical and political factors at the county and intercounty levels, together with explicit specifications of internal dynamics of migation systems. Table 1 displays the results. Model 1 suggests that, holding all other factors constant, a larger difference in political environments between counties predicts less migration between them. Since the political environment is associated with the level of urbanization of a county (Cramer, 2016), Model 2 further includes the dissimilarity of urbanization between counties. From Model 1 to Model 2, the effect size of political dissimilarity becomes modestly smaller, suggesting that the effect of political difference can be partly (but not completely) explained by their difference in the level of urbanization. The smaller BIC of Model 2 further indicates that difference in the level of urbanization is effectively explaining the variation in the magnitude of migration flows. Nonetheless, in Model 2, larger political dissimilarity is still a statistically significant predictor of less migration between counties, offering empirical evidence for Hypothesis 1.1. Holding other factors constant, a pair of counties with 10% larger difference in 2008 voting outcome is expected to have 2.5% (i.e., $[1-\exp(-0.256\times 10\%)]$) fewer migrants than another county pair. Similar to political segmentation, Model 2 also reveals that larger differences in levels of urbanization and racial compositions of two counties predict fewer migrants moving between, holding other factors constant, lending support for Hypotheses 1.2 and 1.3. The VTERGM results do suggest that migration is inhibited between places with dissimilar political contexts, levels of urbanization, and racial compositions.

Table 3.1: Valued TERGMs for Intercounty Migration Flows, 2011-2015

	Model 1		Model 2	
	Estimate	SE	Estimate	SE
Segmented Immobility				
Political dissimilarity	-.368***	.007	-.256***	.007
Rural dissimilarity			-.399***	.004
Racial dissimilarity	-.361***	.006	-.217***	.006
Network Patterns				
Mutuality	.054***	.002	.045***	.002
Log(past migrant flow)	.303***	<.001	.300***	<.001
Waypoint flow	-.014***	.001	-.015***	.001
Destin.log(immigrant inflow)	.062***	.001	.056***	.001
Origin.log(immigrant inflow)	.040***	.001	.035***	.001
Demographics				
Destin.log(population size)	.351***	.002	.351***	.002
Origin.log(population size)	.370***	.002	.373***	.002
Destin.log(population density)	-.077***	.001	-.083***	.001
Origin.log(population density)	-.062***	.001	-.069***	.001
Destin.PSR	.018***	.001	.017***	.001
Origin.PSR	.013***	.001	.013***	.001
Origin.P(White)	(reference group)			
Origin.P(Hispanic)	-.012	.007	-.064***	.007
Origin.P(Black)	.147***	.008	.117***	.008
Origin.P(Asian)	.408***	.020	.467***	.020
Origin.P(other race)	1.031***	.015	.993***	.015
Economics				

Table 3.1: *(continued)* Valued TERGMs for Intercounty Migration Flows, 2011-2015

	Model 1		Model 2	
	Estimate	SE	Estimate	SE
Destin.P(renter)	.405***	.011	.348***	.011
Origin.P(renter)	.507***	.012	.476***	.012
Destin.P(higher education)	.327***	.011	.359***	.011
Origin.P(higher education)	.157***	.012	.153***	.012
Difference.log(housing costs)	-.135***	.004	-.153***	.004
Origin.log(housing costs)	-.248***	.005	-.277***	.005
Difference.P(unemployment)	-1.305***	.040	-1.300***	.040
Origin.P(unemployment)	-3.039***	.052	-3.012***	.052
Geographics				
Log(distance)	-.563***	.001	-.568***	.001
Same state	.501***	.002	.510***	.002
Northeast	(reference group)			
Destin.South	.258***	.003	.253***	.003
Origin.South	.047***	.003	.046***	.003
Destin.West	.384***	.004	.374***	.004
Origin.West	.193***	.004	.184***	.004
Destin.Midwest	.203***	.003	.197***	.003
Origin.Midwest	.085***	.003	.080***	.003
Baseline				
Sum	-1.609***	.040	-1.193***	.040
Nonzero	-13.966***	.028	-13.917***	.028
BIC	2,221,363		2,210,125	

Note: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$ (two-tailed tests).

Table 3.2: Migrant Population Sizes under Observed and Knockout Scenarios

	Total Migrants	Increment in Count and Rate	
Observed	17,176,675		
Remove political segregation	17,965,336	788,661	4.6%
Remove all segmentation	21,741,021	4,564,346	26.6%

To quantify the contribution of segmented effects to immobility, we perform knockout experiments to compute the total migrant population under counterfactual scenarios where these effects are inoperative, and compare that with the observed scenario. Table 3.2 shows that when the political segregation effects on migration flows were knocked out, the expected intercounty migrant population each year would increase by 789 thousand, 4.6% higher than the observed. At the absence of all three segmentation patterns, we would expect to observe 26.6% more internal migrants in the United States, that is 4.56 million more people moving from one county to another each year.[11]

Results of the VTERGMs and knockout experiments together suggest that segmented immobility serves as a critical and substantial social mechanism behind the immobility of the contemporary American society. These social mechanisms may be partly driven by economic forces (although supplementary analysis shows that dual labor and housing markets make little impact on the described segmentation pattern, see Appendix (3.7.4)); it may also reflect people's preference for residing in an environment that is culturally and politically familiar to them. This tendency not only implies social cleavages along party lines, between urban and rural lands, and across communities with varying racial demographics; it could also contribute to a growing geographical segmentation along those lines. As has been known since the classic works of Sakoda and Schelling, even a small preference for homophily can lead to substantial segregation in residential settlement patterns (Fossett, 2006; Sakoda, 1971; Schelling, 1969).

3.5.3 Network Dynamics Influencing Migration Flows

The VTERGMs also consider the network patterns of the migration flow system. That all coefficients are significant in the *Network Patterns* section in Model 2 of Table 3.1 confirms that they play a significant role in determining the directions and magnitudes of intercounty migration flow. In Model 2, the positively significant mutuality term confirms Hypothesis 2.1, that reciprocity is present in the migration-flow networks: a larger flow from county A

---

[11] We note that this conclusion depends on the assumption that the context dissimilarity influences people's decision of whether to migrate or not, and not merely influencing their choice of destination. We would thus not expect this model to accurately predict involuntary migration in response to events like political turmoil or natural disasters, which dominate people's decision of migrating or not under those circumstances. However, these seem unlikely to have been significant drivers of internal migration in the U.S. during the study period. We thank an anonymous reviewer from the *American Sociological Review* for pointing out this assumption.

to B is positively associated with a larger flow from county B to A, holding other effects constant. Joining research on global migration and intercounty migration in U.K. (Ravenstein, 1885; Windzio, 2018), we show that reciprocity is also a network pattern found within U.S. migration. It is interesting to note that some prior studies not observe reciprocity effects in their analyses (Desmarais and Cranmer, 2012b; Windzio et al., 2019); this might come from omission of some regional characteristics that influence the attractiveness of regions to migrants, or their operation of data transformation for the migrant count variable. Future research may replicate the analysis of reciprocity using count-data network models under various social contexts to understand whether reciprocity is a prevalent phenomenon, or can be suppressed by some social forces.

Model 2 also reveals that a larger migration flow during 2006-2010 is significantly associated with a larger migration flow during 2011-2015, even after holding all exogenous and endogenous factors constant. This confirms Hypothesis 2.2 regarding the perpetuation of the migration flow system, showing that migration-facilitating mechanisms offer the system its own momentum, promoting future migration net of exogenous factors such as demographic structures of a region (de Haas, 2010).

The significantly negative coefficient of the flow term indicates a lack of waypoint structures of intercounty migration, refuting Hypothesis 3. The negative waypoint flow effect implies that relatively little migration is proceeding in the chain-like manner such as stepwise and relay migration. After holding other factors constant, counties generally have an imbalance or inequality in the scales of their migration inflows and outflows, either sending many migrants but receiving few, or receiving many migrants but sending few. This may represent emergent attractiveness effects, in which in-migration makes a county seem more attractive to other possible migrants, and out-migration makes a county seem correspondingly less attractive. It may also reflects unobserved heterogeneity in attractiveness arising from other factors; the specification of waypoint flows in the model thus controls for this possible source of autocorrelation, beyond its substantive interest.

Note that the inequality identified by a lack of waypoint flows in this intercounty migration network is different from the inequality captured by an abundance of transitive hierarchy in other cross-national migration networks (e.g., Leal, 2021). The transitive hierarchy requires many waypoints serving as the "mildly structurally attractive position," between the highly and the minimally "structurally attractive positions" (Leal, 2021): 1086). In analogy, that implies a multi-layer hierarchy of the global system with countries positioned in the core, the semi-periphery and the periphery (Wallerstein, 2011). On the contrary, in this network with a lack of waypoint flows, there is an *absence* of semi-periphery areas serving as waypoints between the core and the periphery; in comparison with the international migration system, the U.S. migration system is relatively bipolar, with counties tending to be, *ceteris paribus*, either structurally attractive or unattractive, with few in the middle ground.

The model also examines the relationship between internal and international migration. It shows that larger immigrant inflows from 2011 to 2015 are positively associated with larger intercounty inflows and outflows in the same period. This finding does not correspond to either side in the debate about internal migratory response to immigration, which contends that large immigrant inflows are either associated with small internal migrant inflow and large outflows, or not associated with internal migrant flows. Rather, the results suggest that counties with large immigrant inflows are active in *both* sending and receiving intercounty migrants. Further, the larger coefficient of destination effect than the origin effect suggests that, increasing immigrant inflows to a county is associated with larger increase of internal inflow than internal outflow. In other words, immigration is actually associated with net population increase from internal migration. Overall, the finding shows a common mobility pattern for internal and international migration, wherein counties popular among international immigrants are also popular in receiving and active in sending internal migrants.[12]

3.5.4 Demographic, Economic, and Geographic Determinants of Migration

Alongside segmented immobility and network patterns, the models also consider other factors that could influence intercounty migration. For demographic characteristics, Model 2 confirms findings from spatial econometrics (gravity) models that population sizes in both sending and receiving regions are positively associated with migrant flow (Boyle et al., 2014; Zipf, 1946, 1949). A 10% increase in destination's population size is associated with a 3.4% (i.e., $[1.1^{0.351} - 1]$) increase in the number of migrants, and a 10% increase in origin's population size is associated with a 3.6% (i.e., $[1.1^{0.373} - 1]$) increase in the number of migrants, holding other factors constant. Population density has a significantly negative effect for both the number of in-migrants and out-migrants, holding population size and other factors constant. One possible mechanism is that higher population density leads to larger shares of local connections for their residents (Butts et al., 2012; Hipp et al., 2013; Thomas et al., 2022), where more job transitions and housing transactions can happen locally thanks to these connections, reducing migration across county borders.

With respect to demographic composition, larger migration flows are significantly more likely to be observed between counties with younger populations, in line with the migration schedule literature finding that younger adults are more mobile than older adults (Raymer and Rogers, 2007; Rogers and Castro, 1981). The model also shows that counties with larger shares of Hispanic population tend to send fewer migrants, but counties with larger

---

[12] Since this is an aggregate-level analysis of population flows, the finding does not distinguish the characteristics of internal migrants, such as their race and ethnicity or socioeconomic status. Hence, we do not directly engage with more fine-grained debates about whether immigration deters in-migration and promotes out-migration of certain population categories as predicted by some literature (Frey, 1995a), which requires more detailed data.

shares of Non-Hispanic Black, Non-Hispanic Asian and Other races populations tend to send more intercounty migrants. Note that these effects do not directly describe the mobility of each racial/ethnic population, since they are predicting the magnitude of migration flow for all racial and ethnic populations. Decomposing migration flows into migrants of each racial/ethnic population is necessary to further reveal the variation of mobility between people with different racial/ethnic identities.

Economic covariates in Model 2 show that larger migration flows exist between counties with higher shares of renters and people with college degrees, consistent with previous literature observing that renters and people with higher education credentials are more mobile than their counterparts (Frey, 2009). We also see that larger migration flows happen when the route offers greater declines in housing costs, indicating a tendency of mobility towards cheaper housing (Plantinga et al., 2013). Holding other factors constant, counties with lower housing costs have higher out-migration. This might be due to the better financial conditions renters have in low housing cost areas, enabling them to move and relocate. It is also compatible with previous findings that lower housing equity is associated with higher mobility rates (Coulson and Grieco, 2013). For unemployment rates, the model suggests that the lower the unemployment rate at the origin, and the larger the decline in unemployment rate from origin to destination, the more intercounty migration. These results are compatible with the cost-benefit model of the neoclassical economic theory of migration that population move towards economic opportunities (Todaro, 1976), and that more economic opportunities financing migration makes migration more likely to happen (Massey and Espinosa, 1997). The relational approach employed here enables empirical analysis of the aspiration-ability model (Carling, 2002; Carling and Schewel, 2018), revealing that both the aspiration, as influenced by the relative economic conditions of origin and destination, and the ability, as influenced by the economic conditions of the origin, matter to migration behaviors.

In terms of geographical factors, the model suggests a negative association between distance and number of migrants flowing between two counties, as the gravity model predicts (Zipf 1946, 1949). A 10% increase in distance between two counties is associated with a 5.3% (i.e., $[1 - 1.1^{-0.568}]$) decrease in intercounty migration. Administrative boundaries also influence migration flows; migration flows within the same state are expected to be larger than those across states, holding other factors constant. Additionally, different U.S. regions have varying mobilities. The model indicates that compared to the Northeast, every other region receives and sends more intercounty migrants, *ceteris paribus*. This suggests the existence of some latent characteristics inhibiting the mobility of the Northeast, which deserves more examination in future work.

Lastly, to check the model adequacy, we simulate networks based on Model 2 (the full model) in Table 1 using MCMC algorithms. We then calculate the total in-migrant and out-migrant count for each county, and compare the observed distribution with the simulated distribution. We find that the fitted model recapitulates the county-level migration data (see Appendix (3.7.4)). We also calculate the Pearson's correlation between observed and simulated distributions, which are all above 0.95. We conclude that the model effectively reproduces the quantitative features of observed migration flow networks.

3.6 Discussion and Conclusion

This chapter offers a comprehensive analysis of the intercounty migration structure encompassing not only economic, demographic and geographical factors, but also political, cultural factors and internal dynamics of the migration system. Network models reveal a pattern of segmented immobility in America, in which less migration happens between counties with dissimilar political environments, levels of urbanization, and ethnic/racial compositions. Yet, we do not observe segmentation between internal migrants and international immigrants; rather, the model shows that counties active in receiving many international immigrants are active in both sending and receiving many internal migrants as well. Our analysis also suggests the significance of internal dynamics of the migration flow system; we observe strong patterns of reciprocity and perpetuation, along with a suppression of waypoint structure. These results lend empirical evidence to the systemic theory of migration (Bakewell, 2014; de Haas, 2010; Mabogunje, 1970; Fawcett, 1989), showing that the population flows assemble an interdependent network system that carries its own momentum.

This chapter identifies segmentation as a critical mechanism behind population immobility in the contemporary American society, which could potentially have deterred millions of people from migrating each year, as suggested by the knockout experiments. This finding implies people's tendency of choosing residency in localities that match with their political affiliations and sociocultural attributes, potentially leading to geographical segmentation between people with different political identities (Brown and Enos, 2021) and increasing the homogeneity of their social relations (DiPrete et al., 2011). Such sorting could possibly reinforce political polarization (DellaPosta and Macy, 2015), and can also serve as a mechanism that maintains and even exacerbates residential segregation along other dimensions (Fossett, 2006; Sakoda, 1971; Schelling, 1969). While classic analyses of segregation have focused on local communities within urban areas (Bishop and Cushing, 2009), the effects seen here could potentially contribute to macro-level segmentation across the whole country (Liu et al., 2019). From a migration perspective, although internal migration in the U.S. does not involve border-crossing in international migration or other forms of governmental restrictions (such as the household registration system in China, *hukou*), population movement is never free of constraints. Rather, as our analysis shows, Americans today are separated by the invisible borders and walls standing along the party lines, at the midway between rural and urban landscapes, and over the gap across communities with varying racial demographics.

The analytical framework in this chapter provides an example of structural and systemic analysis of mobility and immobility, broadly defined. The relational approach connects the perspectives of emigration and immigration to examine how characteristics of origin and destination *jointly* influence migration, which enables revelation of the segmented immobility in the U.S. migration system. The formal specification of the interdependence between migration flows under the ERGM framework identifies the structural signature of networks, reflecting the internal dynamics of migration systems. The knockout experiment offers model-based insights into how the system might react to social change. Lastly, leveraging advances in scalable VERGM estimation and simulation allows quantitative analysis of the magnitude of population flows and their determinants in large social systems. The applicability of this framework extends beyond the population movement between geographical areas, encompassing mobility in the occupational system for the study of social stratification and mobility (Cheng and Park, 2020), the exchange of personnel between organizations (Sparrowe and Liden, 1997), and the migration of scholars between institutions and research domains in the sociology of knowledge (Burris, 2004; Gondal, 2018; McMahan and McFarland, 2021).

While our study enables a much richer examination of the mechanisms driving or inhibiting internal migration at a larger scale than what has been possible in extant literature, it is not without its own limitations. First, as a macrosociological study about the "functioning of a social system" (Coleman, 1986): 1312), this chapter informs an aggregate-level social phenomenon, i.e., population immobility. While analysis of the migration flow network facilitates a systemic understanding of migration and its relation to segmentation from a holistic viewpoint, it does not directly describe the patterns of individual migration behavior. Although we can test for the structural signatures of such micro-level processes, unpacking those fine details requires information on decision making and behavior patterns at the individual level. For example, distinguishing stepwise migration and relay migration requires data about the migration trajectories of individual migrants. Studies like this are hence complementary to micro-level analyses (both quantitative and qualitative) that could shed further light on processes at the individual and household levels (e.g., DeLuca et al., 2019; Fitchen, 1994; Lichter et al., 2022; Quillian, 2015). Research that aims to bridge individual behaviors and aggregate social outcomes are deemed to be fruitful, which is still an open problem in sociology, but a promising program to pursue (Cetina and Cicourel, 2014; Coleman, 1986).

Second, since the American Community Survey did not start collecting data until 2005, our analysis only includes migration-flow networks for two time points (2006-2010, 2011-2015). This data limitation prevents us from conducting dynamic analysis about changes in inter-county migration patterns throughout the past decades, and therefore, our findings do not speak directly to the reasons behind the long-term decline of migration. Yet, our identification of drivers and especially inhibitors behind migration flows could serve as a starting point for this inquiry. For example, since political division across geographical areas deters migration, it may be worthwhile for future research to examine how the geography of politics and preference about political homophily have changed over time, and how the evolution of political landscapes and polarization relates to the long-term decline of migration. Studies of the changing patterns of immigrant inflows and the relationship between internal and international migration flows can illuminate the change of population dynamics over time. Integration of knockout experiments via network simulation and historical data about political climate and migration/immigration flows might be one approach to advance the inquiry into the social forces behind the growing immobility in the United States. In addition, future research might also benefit from exploring the changing balance of forces of the competing internal dynamics of the migration system over the past decades. Given that the VTERGM framework we employ here is capable of handling networks with multiple time steps, our analytical framework could be employed for dynamic analysis once migration-flow data for more time points becomes available.

In like vein, the time period we analyzed covers the Great Recession (Grusky et al., 2011). Despite our controls of various economic factors, it is possible that some aspects of our findings may be particular to this period, as economic shocks can influence migration patterns (Monras, 2018; cf. Molloy et al., 2011). Specifically, since economic recession can suppress migration, it is possible that fewer waypoint flows are consequence of the period effect that temporarily suppresses stepwise migration. Nevertheless, the formal expressions of relational linkages and network patterns, and the modeling of migration-flow networks using ERGMs are generally applicable to study migration flows of different periods and regions at different scales. Future research may consider replicate and compare analysis of relational and network patterns of migration flows in different time and space using similar frameworks; they will reveal what patterns are context-specific in certain spatial-temporal settings, and which are generalizable to migration in other societies.

Furthermore, another fruitful direction for future work is to complicate the analysis of internal dynamics of migration system by examining higher-order dependence structure of (valued) networks. One example is network transitivity, a structural feature associated with hierarchy within the migration system (Leal, 2021). We do not observe a strong transitive hierarchical system in the U.S. internal migration system, as indicated by the lack of waypoint flows, *ceteris paribus*.[13] Nevertheless, transitivity is in general a theoretically-interesting dependence structure for study of mobility networks, and should ideally be examined in valued networks so to consider the quantitative feature of migration flows. This requires theoretical and methodological developments in formal specification of dependence terms in the valued network setting, e.g., clarifying the properties of different definitions of transitivity and their relationship to network degeneracy (Krivitsky, 2012). It also demands further advancements in computational methods for valued network models to allow for evaluation

---

[13] As discussed in Hypotheses, both waypoint flow and transitivity are triadic features that concern edge structure in an (i, j, k) triple; waypoint flow captures the“backbone” of flow within the triple $(i \rightarrow j \rightarrow k)$, while transitive triads involve the co-presence of waypoint flow and a direct $i \rightarrow k$ flow. The negative effect for waypoint flow in our models means that triples with strong $i \rightarrow j \rightarrow k$ paths are suppressed, which also necessarily suppresses transitive triples net of other effects in the model. Interestingly, while the waypoint flow (and its binary-network version, two-paths) is a more basic lower-level dependence structure, which carries motivations from social behavior patterns such as those detailed in this chapter, it receives relatively less examination in the network literature. We hope this chapter helps draw more attention to waypoint flow and other triadic network structures of potential substantive importance for flow networks.

of more complicated dependence structures in large networks.

Last but not least, as population immobility has become a long-term phenomenon in the U.S., it poses important questions about its broader social implications. Future research could explore the relationship between geographical mobility and social mobility, and how the divergent geographical mobility patterns across various social groups may influence their life chances and well being. A lack of population exchange, especially between localities with different cultural and political climates, could have ramifications on the social divisions of the country. Two decades ago, Putnam's (2000 *Bowling Alone* embarked the great debates about the "collapse of American communities," marked by the detachment and disengagement of individuals from their communities. Observing the population segmentation and immobility, it raises the question whether we are witnessing the "tribalization of American communities," where local communities diverge in their demographics, culture, and policy, with limited interaction, communication, and cooperation among people and organizations from dissimilar local communities.

In conclusion, grappling with the mobility bias in migration studies, this chapter utilizes migration systems theory and network methods to study the mechanisms behind population immobility in the United States. We identify segmentation as a significant feature of the American migration landscape, which has potentially immobilized millions of intercounty migration each year in the 2010s. The chapter demonstrates how network and simulation methods can contribute to a systemic understanding of mobility and population dynamics. We also call for more theoretical and empirical research about the interrelationships between migration, segregation, and polarization, and how they shape the foundation of social lives in America and beyond.

3.7 Appendix

3.7.1 Descriptive Statistics

Table 3.A1: Descriptive Statistics of Intercounty Migration Flow Networks

	2011-2015	2006-2010
Vertices	3,142	3,142
Edges	274,197	241,526
Density	0.028	0.024
Mean degree[1]	175	154
Total migrants	17,176,675	17,248,855
Mean migrants per county[2]	10,934	10,980
Mean migrants per flow	63	71

Note: 1. The reported degree is the total degree (Freeman degree), which equals the summation of in and out degree (Freeman 1978). For a closed network system, the mean in-degree equals to the mean out-degree. 2. Similarly, the mean migrant per county is the summation of mean in-migrants and mean out-migrants per county (and mean in-migrants equals to mean out-migrants).

Table 3.A2: Descriptive Statistics of Vertex/County Covariates

	Mean	Str Dev	Source
Networks			
Immigrant inflow	593.9	2744.3	ACS2011-2015
Politics			
Democrat poll (%)	41.6	13.9	Presidential Election Returns 2008
Demographics			
Population size	98,262.0	312,946.7	Census2010
Population density (/km^2)	100.0	665.8	Census2010
Potential Support Ratio (PSR)	4.4	1.5	Census2010
White (%)	78.3	19.9	Census2010
Black (%)	8.7	14.4	Census2010
Hispanic (%)	8.3	13.2	Census2010
Asian (%)	1.1	2.5	Census2010
Other race (%)	3.5	8.3	Census2010
Economics			
Renter (%)	27.8	7.7	Census2010
Higher education (%)	19.0	8.7	ACS2006-2010
Unemployment (%)	7.5	3.4	ACS2006-2010
Housing costs ($/month)	707.5	272.3	ACS2006-2010
Geographics			
Rural population (%)	58.7	31.5	Census 2010
Northeast (%)	6.8		U.S. Census Bureau
South (%)	45.3		U.S. Census Bureau
West (%)	14.3		U.S. Census Bureau
Midwest (%)	33.6		U.S. Census Bureau

Table 3.A3: Descriptive Statistics of Dyadic/County-Pair Covariates

	Mean	Str Dev	Source
Dissimilarity			
Political dissimilarity (%)	15.6	11.8	Presidential Election Returns 2008
Rural dissimilarity (%)	36.1	26.1	Census2010
Racial dissimilarity (%)	24.8	20.0	Census2010
Geographics			
Distance (km)	1,439.5	961.3	Census2010

3.7.2 Model Setup and Computation

Model Setup

The exponential-family random graph models, formally speaking, specify the probability of observing a specific network configuration, out of all possible configurations, given the vertex/node set and covariates (Hunter et al. 2008b). In our context, this constitutes the probability of obtaining the observed migration flow structure among the 3,142 U.S. counties, versus the other structures that could have been observed. Formally, our model family is defined by

$$\Pr(Y^t = y^t) = \frac{h(y^t) \exp\left(\theta^T g\left(x, y^t, y_{t-1}^{t-k}\right)\right)}{\kappa(\theta)} \tag{3.1}$$

where Y^t is the valued network structure at time t (with observed value y^t), $g(x, y^t, y_{t-1}^{t-k})$ is a vector of sufficient statistics, and θ is a vector of coefficients. The model statistics are functions of (valued) graph structure, together with covariates (x) and past history (represented by the lagged variable $y_{t-1}^{t-k} = (y^{t-1}, \ldots, y^{t-k})$). The exogenous covariates are characteristics not determined by network configurations, including demographic, economic, political, and geographical attributes of individual counties or county pairs. g may also include statistics relating to endogenous structure, such as reciprocity and the waypoint structure. The use of lagged network predictors allows us to incorporate predictors related to network dynamics (Hanneke et al. 2010) in analogy to a lagged multivariate regression model.

In equation 3.1, $\kappa(\theta)$ is the normalizing factor for the exponential family model, the summation of the quantity of the numerator for all possible network configurations:

$$\kappa(\theta) = \sum_{y' \in \mathfrak{Y}} h(y') \exp(\theta^T g(x, y', y_{t-1}^{t-k})$$

where $\mathfrak{Y}$ is the set of all possible network configurations. The *reference measure*, $h(y)$, is a function that determines the baseline distribution to which the model converges as $\theta \to 0$. This is particularly important for valued ERGMs (Krivitsky 2012), as different choices of h can lead to very different distributions of edge values (just as different choices of reference measures for one-dimensional exponential families differentiate e.g. the binomial, geometric, and Poisson random variables in generalized linear models). Here, we use the form

$$h(y) = 1/ \prod_{(i,j)\in \mathbb{Y}} y_{i,j}!$$

where $\mathbb{Y}$ is the set of all dyads, and $y_{i,j}$ is the value of the edge from county i to j. This leads to a model where edge values are Poisson-like in the absence of dependence effects; just as a conventional ERGM can be viewed as a network logistic regression extended to incorporate dependence, this family can be viewed as a network Poisson regression with dependence among the edge variables. (Note, however, that the addition of dependence terms renders the model non-Poissonian, and a model of this form can accommodate e.g., over-dispersion and/or zero-inflation by appropriate choice of terms, as is done here.)

Computation

Estimation for ERGMs can be computationally demanding, since the normalizing factor κ required for likelihood computation is too expensive to compute directly. This is especially the case for large and high-variance valued networks such as intercounty migration-flow networks, where the number of edge variables is extremely large (nearly ten million), and the edge value ranges from zero to tens of thousands. Researchers typically use Markov Chain Maximum Likelihood Estimation (MCMLE) for ERGM parameter estimation when modeling binary networks, which uses a Markov Chain Monte Carlo (MCMC) algorithm to produce importance samples of network structures that are subsequently used for approximate maximum likelihood inference (Geyer and Thompson 1992; Snijders 2002). Unfortunately, current MCMLE implementations for valued networks (including both Geyer-Thompson-Hummel and stochastic approximation methods) do not scale well enough to be feasible for networks studied here, with such large network size and edge variance, as the space of potential random networks is too large for the MCMC sampler to converge in feasible time. Instead, we employ an efficient parallelizable subsampled regularized Maximum Pseudo-Likelihood Estimation (Huang and Butts 2024b). Maximum Pseudo-Likelihood Estimation (MPLE) is a classical computational method for ERGMs (Besag 1974; Strauss and Ikeda 1990; Wasserman and Pattison 1996), which continues to be employed as an efficient strategy for computationally intensive ERGMs (An 2016; Schmid and Desmarais 2017). MPLE approximates the network likelihood function as the product of the conditional probability functions of each edge; this approximation becomes exact in the limit of weak edgewise dependence. This allows parallel computing and subsampling of edges to reduce computation time. While MPLE performance is not always equal to MCMLE performance for binary ERGMs, simulation studies find that for count-valued ERGMs with large edge variance, MPLE can match or even outperform MCMLE with reliable and nearly unbiased estimates, while its computational cost is substantially smaller (Huang and Butts 2024b). We further incorporate an L2 (ridge) regularizer to our likelihood function, making it Regularized MPLE, following the suggestion of using regularization to improve MPLE performance in the literature (van Duijn et al. 2009).

Since calculating the likelihood function of valued TERGMs based on all 9.87 million dyads is computationally infeasible, we use the subsampling implementation of MPLE to approximate the likelihood function using one million dyads. The sampled dyads were selected using Tie-No-Tie (TNT) sampling, which puts equal weight on selecting zero-valued dyads and nonzero-valued dyads. TNT is an efficient sampling method for ERGMs when the network density is low, as in the present case where the majority of county pairs have no migrants. This results in including all the 274 thousand directed county pairs with at least one migrant, and randomly sampling the other 726 thousand directed county pairs without any migrant; this stratification is then corrected for using inverse probability weighting in the likelihood calculation.[14] L2 regularization (parameter=0.01) was incorporated in the likelihood function to avoid the convex hull problem and improve estimation performance (van Duijn et al. 2009).

---

[14] The standard errors calculated from the sampling method are expected to be slightly larger than using the full network data (Huang and Butts 2024b), as the former uses fewer observations, making our inference more conservative. Since our main findings are statistically significant at the level of 0.001 under a conservative protocol, we expect the effects to be significant if the computational constraints had allowed us to use the full network for calculation.

3.7.3 Notes about *In Silico* Experiments

In silico experiments are thought experiments demonstrated by computer simulation. Sometimes based on an empirically calibrated model, the experiment constructs counterfactual scenarios that are not observable. This offers great flexibility for researchers to posit various scenarios and gain insights. On the other hands, it requires users to ponder the study design and the assumptions made.

There are different ways of probing the model via *in silico* experiments. This chapter focuses on the knockout experiment, which examines the scenario where certain effects were eliminated. Adopting the terminology from Butts (2023), one can also *knock up* (enhance) certain effects or *knock down* (reduce, but not eliminate) them. One may also *knock in* effects by including effects not described in the model. Moreover, one can perform context-shifting experiments, altering the covariate value to examine counterfactual social contexts. As we will discuss below, although knockout and context-shifting experiments answer distinct questions, the former has some intrinsic connection to the latter.

Focusing on the knockout experiment, it is critical to consider how the model specifies the baseline scenario. When knocking out an effect, we assume the system returns to its baseline state, rather than the observed one shaped by the effect of interest. The baseline state can be clear and obvious in some cases; for instance, when knocking out an intervention (e.g., vaccination, financial aid), we return to a state where this intervention does not exist. However, the baseline state can sometimes be implicit. Taking the example in this chapter about the effect of political climate on migration rate, what is the baseline migration rate when political climate no longer matters? According to our model, political dissimilarity between counties has a negative effect on their migration rate: $E(Y_{ij}) = f^{-1}(\beta_0 + \beta_1 \cdot D_{ij})$.[15]

---

[15] For simplicity, we focus on a specific migrant count from county i to j, Y_{ij}, and use a generalized linear model description, with $f(\cdot)$ serving as the link function, and $E(\cdot)$ as expectation. $D_{ij} = |X_i - X_j|$ is the dissimilarity in political climate X, and β_1 is its association with migrant count.β_0 is intercept, or for multi-covariate model is the summation of the intercept and the rest of model parameters that will be held

We knock out this effect by assuming no influence from political dissimilarity, setting the coefficient to zero: $\beta_1 \doteq 0$. This gives us $E(Y^a_{ij}) = f^{-1}(\beta_0) = E(Y_{ij}|D_{ij} = 0)$, which is equivalent to the scenario where the political dissimilarity between these two counties is zero. So this model specifies a baseline state where no political dissimilarity occurs. In other words, the model believes that when political climate does not influence migration, it is as if there exists no political dissimilarity; the observed political dissimilarity suppresses migration, so when the suppressive effect is knocked out, migration rate is elevated.

Alternatively, one can specify an opposite model believing that when political climate does not influence migration, it is as if there exists no political similarity ($S_{ij} \doteq 100\% - D_{ij} = 0$) between two counties; that is to have one county with 0% Democratic voter and the other with 100%. Then, the observed political similarity boosts the migration rate, and knocking out this boosting effect reduces migration rate. Formally speaking, the model is $E(Y_{ij})) = f^{-1}(\beta'_0 + \beta'_1 \cdot S_{ij})$, and the knockout scenario is $E(Y^b_{ij}) = f^{-1}(\beta'_0) = E(Y_{ij}|S_{ij} = 0) = E(Y_{ij}|D_{ij} = 100\%)$. The two models offer distinct knockout results because they have distinct assumptions about the baseline scenarios. These assumptions are well defined in the model, though one should examine their sensibility before performing *in silico* experiments.

Exploiting the idea that knocking out an effect means returning the system to the baseline state specified by the model, one can also interpret the knockout experiment as a result of altering the covariate value. Mathematically, since the effect is expressed as $\beta_1 \cdot D_{ij}$, it is equivalent to eliminate the effect by having $\beta_1 \doteq 0$ or by having $D_{ij} \doteq 0$. In the previous example, knocking out dissimilarity effect generates a counterfactual migration system equivalent to the one generated by assuming no dissimilarity exists, a world where all U.S. counties have the same political climate. So the knockout experiment performed in this chapter simultaneously answer two distinct questions: 1) what is the migration rate when political dissimilarity does not influence migration, and 2) what is the migration rate

---

constant in experiment $\beta^T g$).

when no political dissimilarity exist among counties.

There are two caveats when performing and interpreting *in silico* experiments by altering covariate values. First, altering covariate values and knocking out effects (coefficients) are not always equivalent. The example in the next chapter demonstrates this point: when one covariate (e.g., housing costs) influence the outcome (e.g., migration rates) in multiple ways (as expressed by multiple terms in the model), having one mechanism turned off by setting its coefficient to zero is not equal to altering its covariate, because the latter will impact the outcome via other specified mechanisms. Second, it is also necessary to consider the sensibility of the counterfactual world specified by the altered covariate. For instance, in the alternative similarity model specified above, it is problematic to assume a world where all pairs of counties having 100% dissimilarity.[16]

It is also worth noting that, although we use reverse coding (similarity v.s. dissimilarity) as a way of demonstrating different possible baseline assumptions, what really matters here is not the choice of coding, but more fundamentally the counterfactual scenario one is interested in. For instance, knocking out dissimilarity ($D_{ij} \doteq 0$) is equivalent to knocking up similarity to its maximum ($S_{ij} \doteq 100\%$); we can answer the same question under different coding choices.

Lastly, while in this chapter the outcome of interest of the experiment coincides with the outcome of interest of the empirically calibrated model, namely the migrant count, this need not to be the case. Researchers can simulate counterfactual scenarios and examine outcomes different from that of the model. As demonstrated in the next chapter, for example, we can use a similar model to simulate and examine the asymmetry in migration flows in revelation of relative migration gain and loss among U.S. states. The reason of performing simulation is that the macro-level outcome is not always a simple summation of its micro-level processes in complex systems, such as networks (see also Duxbury, 2023); model-based simulation

---

[16] This is because that will require all dyads to have one node's covariate value=0, the other=1. In that case, for a closed system, we will always find dyads connecting two nodes with both covariate values=0 or =1.

takes into account the spillover of local effects and other types of endogenous processes exemplified by the internal dynamics of migration systems discussed in this chapter. As long as the outcome of interest is reasonably calibrated by the empirical model (i.e., the model can adequately captures this feature), it is suitable to serve as a target of examination.

3.7.4 Supplementary Analysis

This section considers a potential confounding social process that could be correlated with our main findings.[17] Since people with and without college degrees may respond to different labor market segments under the dual labor market theory (Piore 2018), it is possible that college graduates are more likely to migrate to another place with concentrated college graduates, where there tends to be more job opportunities that they fit into; the same logic holds for people without college degrees. This means that larger migration can occur between counties with similar proportion of college graduates, because of the match in labor market supply and demand. Similarly, we may observe larger migration flow between counties with similar proportion of renters. The proportions of college graduates and renters of a county correlate with its political environment, urbanization level, and racial compositions, so it is desirable to examine how the hypothesized processes may be associated with our major findings. We add covariates of the absolute differences in college graduate share and renter share in Model 3 of Table S4, displayed alongside with the model in the main text (Model 2). The results show that in Model 3, the coefficients representing segmented immobility patterns all remain statistically significant, with little change in its scale for the political covariate, a modest decrease for the urbanization covariate, and a modest increase for the racial composition covariate. The analysis shows that this social process does not confound our main findings.

Table 3.A4: Valued TERGMs for Intercounty Migration Flows, 2011-2015

	Model 2		Model 3	
	Estimate	SE	Estimate	SE
Segmented Immobility				
Political dissimilarity	-.256***	.007	-.257***	.008
Rural dissimilarity	-.399***	.004	-.298***	.004
Racial dissimilarity	-.217***	.006	-.246***	.006
Network Patterns				

---

[17] We thank an anonymous reviewer from the *American Sociological Review* for suggesting this possible social process.

Table 3.A4: *(continued)* Valued TERGMs for Intercounty Migration Flows, 2011-2015

	Model 2		Model 3	
	Estimate	SE	Estimate	SE
Reciprocity	.045***	.002	.036***	.002
Waypoint flow	-.015***	.001	-.016***	.001
Log(past migrant flow)	.300***	<.001	.296***	<.001
Destin.log(immigrant inflow)	.056***	.001	.056***	.001
Origin.log(immigrant inflow)	.035***	.001	.034***	.001
Demographics				
Destin.log(population size)	.351***	.002	.355***	.002
Origin.log(population size)	.373***	.002	.376***	.002
Destin.log(population density)	-.083***	.001	-.086***	.001
Origin.log(population density)	-.069***	.001	-.073***	.001
Destin.PSR	.017***	.001	.020***	.001
Origin.PSR	.013***	.001	.016***	.001
Origin.P(White)	(reference group)			
Origin.P(Hispanic)	-.064***	.007	-.029***	.007
Origin.P(Black)	.117***	.008	.106***	.008
Origin.P(Asian)	.467***	.02	.592***	.02
Origin.P(other race)	.993***	.015	.972***	.016
Economics				
Destin.P(renter)	.348***	.011	.318***	.012
Origin.P(renter)	.476***	.012	.409***	.012
Dissimilarity.P(renter)			.309***	.010
Destin.P(higher education)	.359***	.011	.624***	.012
Origin.P(higher education)	.153***	.012	.440***	.013
Dissimilarity.P(higher education)			-1.051***	.009
Difference.P(unemployment)	-1.3***	.040	-.969***	.040
Origin.P(unemployment)	-3.012***	.052	-2.297***	.052
Geographics				
Log(distance)	-.153***	.004	-.171***	.004
Same state	-.277***	.005	-.327***	.005
Northeast	(reference group)			
Destin.South	-.568***	.001	-.572***	.001
Origin.South	.510***	.002	.518***	.002
Destin.West	.253***	.003	.255***	.003
Origin.West	.046***	.003	.053***	.003
Destin.Midwest	.374***	.004	.379***	.004
Origin.Midwest	.184***	.004	.191***	.004
Baseline				
Sum	.197***	.003	.197***	.003
Nonzero	.080***	.003	.086***	.003
BIC	2,210,125		2,196,016	

Note: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$ (two-tailed tests).

3.7.5 Model-Adequacy Checks

We develop a procedure to evaluate the model adequacy for VERGMs emulating that for binary ERGMs (Hunter et al. 2008a). Based on Model 2 in Table 1 (the full model), we simulate 100 networks using MCMC algorithms using `ergm.count` package in `R`, and calculate the distributions of in- and out-volumes of edge values, i.e. the total in-migrants and out-migrants, for each node (county). We plot the distributions of in- and out-volumes by county against their observed statistics, and calculate the correlations between them. An adequate model specification is expected to reproduce the observed statistics with distribution centered on and close to observed values, and have high correlations between observed and simulated statistics.

Figure 3.A1 and 3.A2 visualize the simulated distributions against the observed statistics of in- and out-migrants, respectively. To make the comparison more viewable than stacking 3,142 counties in one plot, we sample four states from the West, the Midwest, the Northeast, and the South; the y-axis is re-scaled by taking logarithm considering their high skewness. Noticeably, the simulated interval of each county looks very narrow compared to the full range of in- and out-migrant count in each state, especially compared to the common goodness-of-fit plots for binary ERGMs. This indicates that the support of the valued ERGMs is tremendously huge, suggesting that reproducing the migrant distribution is a difficult task. The task is also more difficult because we compare simulated distribution with observed statistics for each county, rather than comparing network-level statistics for an unlabelled graph, which does not consider whether the simulated statistics match the observed at the node level. Nevertheless, the figures show that the simulated distribution is centered around and close to the observed statistics, and the Pearson's Correlations are all above 95%. We conclude that the model effectively reproduces the migrant distribution across counties, and accounts for its variation.

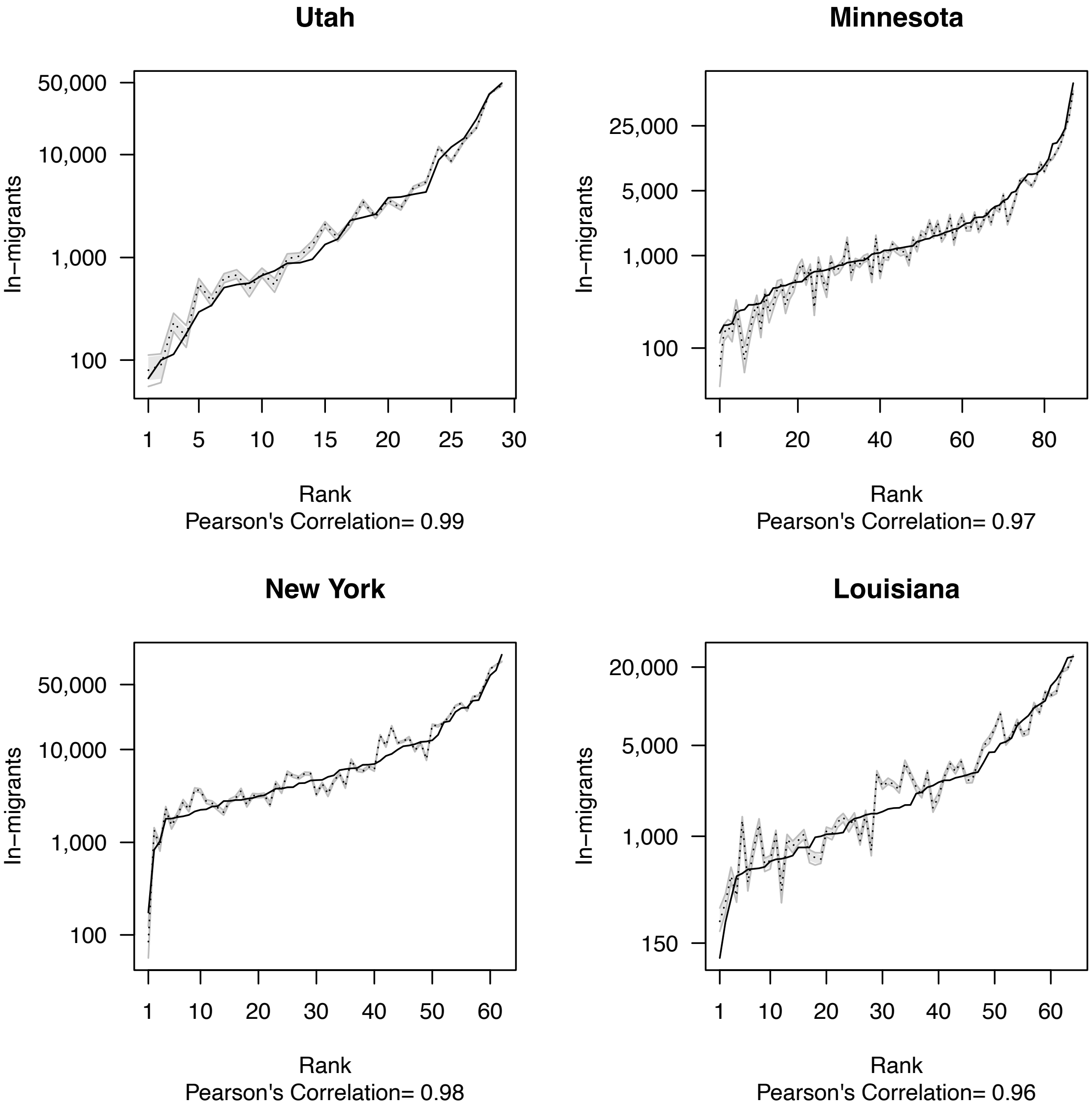


Figure 3.A1: Model-adequacy plots (involume)

Note: The solid dark lines indicates the observed statistics, the dotted dark lines is the median of simulated distributions, grey lines are the min and max values of simulated distributions and the grey polygon is the 95% simulation interval of the distribution.

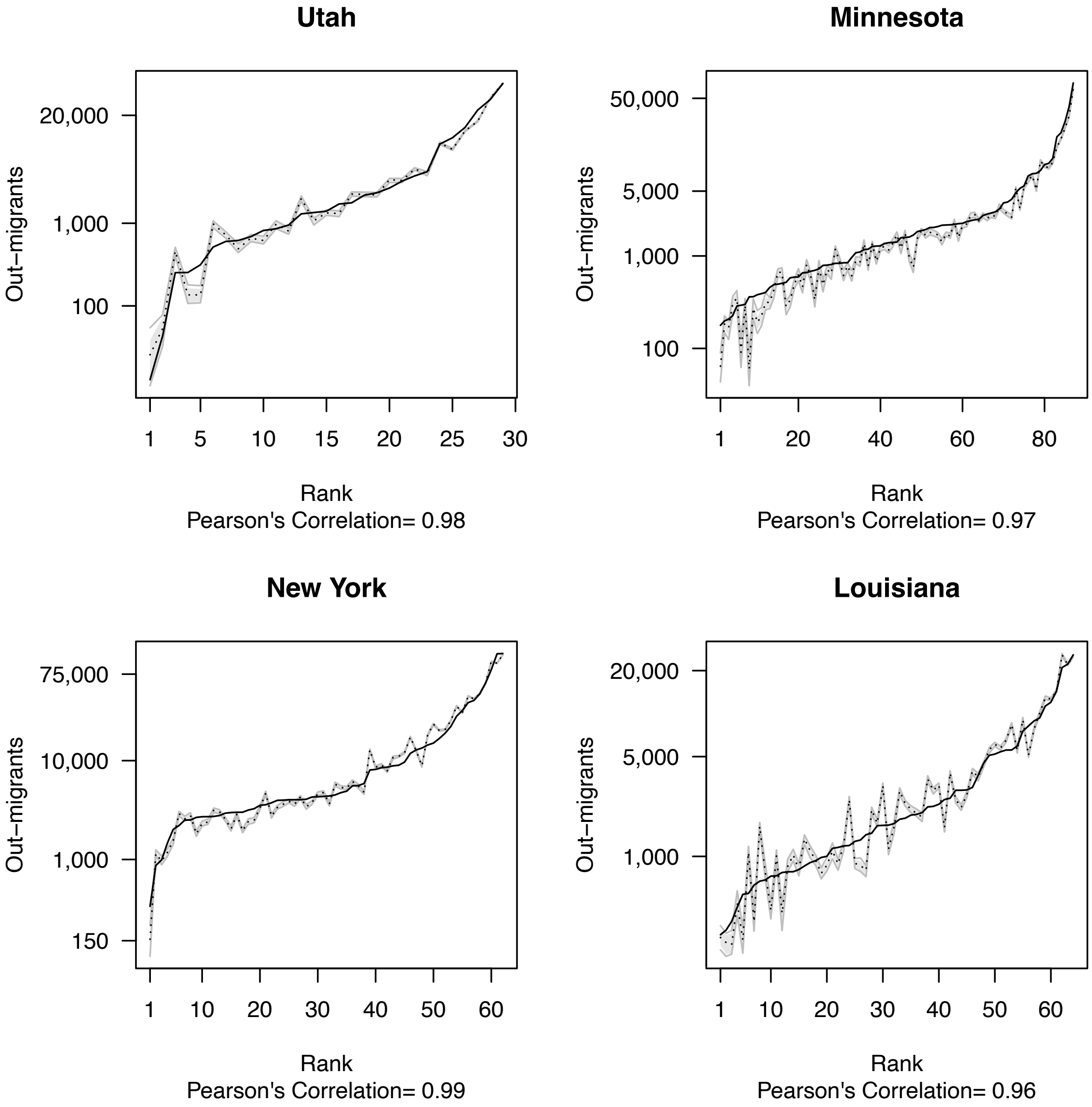


Figure 3.A2: Model-adequacy plots (outvolume)

Chapter 4

California Exodus? A Network Model of Population Redistribution in the United States

4.1 Introduction

The "California Exodus" - a putative phenomenon in which large numbers of individuals are allegedly leaving California and migrating to other U.S. states, has become an increasingly common topic in public discourse surrounding migration and policy in the United States (e.g. Bahnsen, 2021; Beam, 2021; Dorsey, 2021; Hiltzik, 2020; Song, 2021). Popularized within conservative media circles (Bahnsen, 2021; Dorsey, 2021), the notion of a "California Exodus" serves as a focal point for a political narrative in which the state of California exemplifies the failure of the ruling Democratic party governance, and its associated social and policy regimes. Despite this politicized narrative, the net loss of California population via domestic migration *is* a long-term phenomenon, well-documented in demographic data. Nor is this a recent development: contrary to popular impression, California's net migration rate has been negative since 1989 (Hiltzik, 2020). The migration pattern of America's most populous state illuminates important trends of population redistribution in the United States, and could potentially shift the country's economic and political landscape. Historically, internal migration has played a key role in shaping the spatial distribution of population, with the most well-known and general example being urbanization (Ravenstein, 1885). In the U.S., internal migration has also played a critical role in its demographic change, including the great migration of African Americans from the South to the North (Tolnay, 2003), the westward shift of population towards the Pacific coast (Plane, 1999), and the ex-urbanization process (Plane et al., 2005).

Yet, compared to its intense treatment in popular discourse, the California Exodus as a real and persistent (if less dramatic) phenomenon receives scant attention in scientific research (c.f. Henrie and Plane, 2008). Arguably, this may be in part due to the difficulty of modeling the complexity of internal migration systems, which requires incorporating a wide range of factors influencing migration. Moreover, as migration systems theory contends (Bakewell, 2014; de Haas, 2010; Mabogunje, 1970), the migration system has endogenous feedback mechanisms, where migration flows are interdependent to each other. This further complicates mathematical models of migration flows - and their calibration to empirical data - requiring them to account for the autocorrelation structure of the system.

In this chapter, we use recently developed generative network models of the internal migration system in the U.S. to help unravel the mechanisms sustaining the California Exodus, with an eye to identifying factors that may or may not contribute to this feature of the current U.S. migration system. We model the U.S. internal migration system as a network comprising counties (nodes/vertices) and migration flows between each directed pairs of counties (edges). Compared to the conventional approach that considers places as analytical units, the relational approach takes migration flows between places as units of analysis, which allows us to capture how the characteristics of origin and destination *jointly* influence their migration flows, such as the difference in political climates and costs of living. The systemic view also considers the endogenous feedback mechanism of the migration system (de Haas, 2010), reflected by the interdependence among migration flows, which gives the system its own momentum, strengthening or ameliorating the exogenous effects from the economic or political landscapes. This is achieved by specifying the network dependence structure, which accounts for the autocorrelation pattern among migration flows. The network models thus can reveal how demographic, economic, political, and geographical characteristics, together with the endogenous feedback mechanisms, shape the direction and magnitude of internal migration flows in the United States.

While computational and statistical constraints have traditionally limited network models of migration to dichotomous or coarsened representations of migration flows, we use recent innovations in valued exponential-family random graph modeling (Valued ERGMs or VERGMs) to estimate a fully quantitative model of interdependent U.S. migration flows at the county level. Motivated by the popular discourse surrounding the California Exodus and existing theoretical and empirical research regarding U.S. internal migration, we focus on four potential social forces that contribute to population redistribution. They include costs of living, political environments, levels of urbanization, and racial demographics.

This relational view offers new opportunities for insight, but also poses challenges. For instance, interpretation of the relationship between nodal or dyadic attributes' impacts on migration (i.e., covariate effects) can be complex, as such relationships are subject to both the origin's and the destination's attribute values, and they can take various functional forms. Further, the superposition of forms from multiple effects can make the model difficult to interpret. Such complexities reflect the inherent challenges of capturing an interactive system in quantitative detail, and are thus not unique to migration systems, but are particularly acute when considering networks with valued edges. We here propose a visualization protocol that showcases how multiple mechanisms involving origin and destination attributes combine to influence the expected number of migrants between origin and destination regions. We utilize this approach to display how the political, racial, rurality, and housing covariates influence the predicted migration flow intensity across different scenarios, offering a quantitative exploration the impact of dyadic factors on migration.

Another advantage of the VERGM approach is that it offers *generative models,* which can themselves be used to probe the effects of inferred or hypothetical mechanisms beyond the dyadic level. Here, we use our empirically-calibrated migration model to perform *in-silico knockout experiments* to investigate how various social, economic, and demographic mechanisms contribute to observed patterns of population redistribution - including, specifically, maintenance of the California Exodus. These knockout experiments simulate migration flow networks under counterfactual scenarios where certain social effects are inoperative (Huang and Butts, 2023). Comparing the extent of California's relative net migration loss in the knockout scenarios with that in the observed scenario offers quantitative insights about the impacts of social effects on the pattern of population redistribution.

The remainder of the chapter proceeds as follows. We begin in Section 4.2 with a brief review of different approaches to modeling migration systems, and the extant empirical research that motivates our hypotheses regarding population redistribution in the U.S. Section 4.3 describes the data and variables we use, the model setup including the functional form specification, derivation of the visualization protocol, and the knockout experiment procedure. In Section 4.4, we first offer an overview of the population redistribution pattern in the United States, and the pattern of net migration exchange between U.S. states. We then report our findings regarding the drivers of migration patterns from the ERGM analysis, and show how contributing effects can be visualized. The section concludes with results from knockout experiments. The last section summarizes our empirical findings, our contributions to the mathematical modeling of complex social systems, and some directions for future work.

4.2 Background

4.2.1 Modeling Migration Systems

Migration flows among geographical areas form a complex system, a perspective that has received extensive theoretical discussion in migration studies, in the school of *Migration Systems Theory* (MST, Bakewell et al., 2016a; DeWaard and Ha, 2019; Fawcett, 1989; Kritz et al., 1992; Mabogunje, 1970). MST introduces two insights regarding migration. First, a migration system consists of flows of people, goods, information, cultures, and other institutions that interact with each other (Bakewell, 2014). This suggests that understanding migration processes demands a comprehensive survey of various factors and mechanisms, incorporating economic, political, geographical, and demographic analyses. Second, MST emphasizes the *interdependent* feature of migration systems, reflected in their conceptualization of "internal dynamics" (de Haas, 2010) or "feedback mechanisms" (Bakewell, 2014). The central idea is that there exist endogenous processes, where change in one part of the system can diffuse and alter other parts, creating a systemic momentum. This means that migration flows are correlated to each other. For instance, the migration flow from Seattle to Chicago is associated with the reverse flow from Chicago to Seattle, partly because migrants can carry social connections and useful information from their origin to their destination, motivating and facilitating migration in the reverse direction. Such interdependence among migration flows requires mathematical models of migration to account for the autocorrelation among their observations, and ideally, to also formally and explicitly describe the structure of the dependence.

Researchers have developed various methods to model migration across disciplines including econometrics, geography, statistics, and sociology. A convenient and widely used approach is to treat migration as a feature of areal units, analyzing how the characteristics of a place are associated with marginal migration rates into and out of it (e.g., Partridge et al., 2012;

Treyz et al., 1993). This approach has offered many useful insights and serves as a powerful framework for building predictive models of demographic change (Azose and Raftery, 2015, 2018). Methodologically, techniques to account for the autocorrelation in this data structure (areal/lattice data) are well developed in spatial statistics (Banerjee et al., 2014). However, migration is by nature a *relational* process between two places: origin and destination. The above approach by construction marginalizes migration either from an origin perspective or a destination perspective (or condenses both), obscuring how origin and destinations jointly and interactively shape the migration flows between them; such interactions are known to be of considerable importance, as articulated in the classical "push-pull" factor model (Lee, 1966) of migration. From a network analytic perspective, such models are equivalent to modeling the migration network purely in terms of expected outdegree and indegree effects (sometimes called *expansiveness* and *popularity* in the ERGM literature (Holland and Leinhardt, 1981)). Although simple, such models are very constraining - they are essentially similar to a single-dimensional singular value decomposition (SVD) approximation of the adjacency matrix - and are limited in their ability to represent complex structure.

A second model family is the so-called "gravity model" (widely used in spatial econometrics), whose unit of analysis is no longer a geographical area but flow within an ordered pair of geographical areas (i.e., an *edge variable*). The original idea of this model family is that the extent of migration flow from origin i to destination j (M_{ij}) is positively associated with population sizes in origin and destination (P_i, P_j) and negatively associated with the distance between (D_{ij}), with the decay usually posited to follow a power law (Zipf, 1946), thus superficially resembling gravitational attraction.[1] Formally, this family is written as

$$M_{ij} \approx C \cdot \frac{P_i^{\alpha} \cdot P_j^{\beta}}{D_{ij}^{\gamma}},$$

where C, α, β, γ are positive parameters. Although nonlinear on its original scale, the power

---

[1] This formulation is also used to describe other types of spatial interactions such as international trade; see e.g. the review of Anderson (2011).

law model is intrinsically linear, as shown via the log space representation

$$\log M_{ij} = \mu + \alpha \log(P_i) + \beta \log(P_j) - \gamma \log(D_{ij}) + \epsilon_{ij}.$$

where $\mu = \log C$ and log error ϵ_{ij} are unknowns. Factors other than distance and population size may be incorporated by choosing a suitable regression form for μ. The linear form has facilitated further elaboration, e.g. using a GLM structure to capture discrete outcomes (e.g., Biagi et al., 2011). Although the gravity model does not provide a means of specifying dependence among flows, some extensions in this direction have been proposed (see reviews by Patuelli, 2016; Poot et al., 2016).

The gravity models have always been in close relationship with network models, with abundant shared knowledge and mutual development. Fundamentally, gravity models constitute a particular class of network regression models (albeit not necessarily OLS network regression, e.g. Krackhardt (1988)), a very flexible and successful family. Substantively, the functional form of the gravity model arises naturally as a model for *tie* (or interaction) *volumes* between regions under power-law spatial interaction functions, a widely observed functional form for interaction probabilities at the individual level (Butts and Acton, 2011); this, along with the strongly predictive power of distance itself for social networks (Butts, 2003), has been argued to provide a mechanistic explanation for why aggregate interactions are often well-approximated by gravity models (Almquist and Butts, 2015). The identification of gravity models with network regression also points to their limitations: while very flexible in specifying relationships between covariates and tie values, network regression models do not specify dependence among edge variables. While workarounds such as quadratic assignment procedure (QAP) tests (Dekker et al., 2007; Krackhardt, 1988) can provide statistical answers that are robust to dependence effects, parameterization and/or generation of networks with dependence requires other approaches.

The specification of models for networks with complex dependence among edge variables is a major concern of work on exponential-family random graph models, which we discuss in detail in Section 4.3.2. ERGMs provide a rich language for specifying interdependencies among edges, as well as associated statistical theory and methodology for inferring such dependencies from observed network data. Importantly, ERGMs are *generative* - i.e., they provide a full probability model for the target network, and thus can be used for hypothetical realizations of an inferred data generating process. This makes them especially well-suited to mechanistic investigation using approaches such as *in silico "knockout" experiments* and other computational techniques. The increasing availability of scalable and valued-data ERGMs opens the door to modeling migration systems in a substantively-richer and more statistically-rigorous way.

As noted, one advantage that ERGMs have is the ability to explicitly and formally describe the interdependence of edges within networks. In connection with MST, researchers have utilized this feature to formalize and test the patterns and mechanisms of the endogenous feedback processes in migration systems (Huang and Butts, 2023; Leal, 2021; Windzio et al., 2019). Specifying dependence structure can also improve statistical inference. The autocorrelation among migration flows can not only introduce associations in residuals, but may as well impose more general autoregressive structure. In this case, methods that focus on correcting for correlation in the residuals (e.g., QAP) could be insufficient, running the risk of failing to account for the impact of endogenous factors on covariate effects.

Likewise, the generative aspects of ERGMs are particularly relevant in the context of studying migration systems. The ability to simulate from empirically calibrated or *a priori* models allows researchers to extrapolate models across spatial and temporal contexts and even investigate counterfactual scenarios. Although there is work in this direction (including applications to the study of migration systems (Huang and Butts, 2023)), it is arguably an under-appreciated property of this model family, which has been mostly employed as a tool for hypothesis testing. This chapter aims to exploit the generative capacity of ERGMs to quantify the contribution of various drivers of population redistribution to the California Exodus.

Despite these advantages, using ERGMs to study migration systems poses a number of challenges. First, it can be computationally intensive to fit (and sometimes to simulate draws from) such models, since closed-form (or even directly computable) expressions for the likelihood are not attainable except in special circumstances. Moreover, generative models for valued/weighted networks are less developed than binary networks, in terms of formal specifications of dependence structures, theoretical justifications of those specifications, and efficient computational tools; this means that researchers sometimes have to dichotomize migration flows, losing critical information about the scale of migration flows. While it is not the focus of the chapter to advanced generative models for valued/weighted networks, we employ recent advances in this area to offer a quantitative understanding of the population redistribution pattern within the United States.

Moving beyond ERGMs *per se*, a general challenge in modeling relational data such as migration system data is understanding the combined effects of multiple influences, since prediction of a specific migration flow usually involves attributes from different sources (e.g., origin and destination) that can be combined in different ways. The usual approach of interpreting coefficients separately under the *ceteris paribus* condition is often unhelpful here, as these covariates are intrinsically inter-related. For example, often it is substantively natural to include covariate factors (e.g., housing costs) of origin, destination, and their absolute difference, where the last term can no longer be interpreted only as a dissimilarity measure since the statistic is fixed once we hold constant the origin and destination covariates. This chapter tackles this problem by introducing a visualization protocol that helps interpret the multiplex of inter-correlated functional forms that is common in relational data analysis.

4.2.2 Drivers of Population Redistribution

This section examines possible drivers of population redistribution, with an empirical focus on the case of California Exodus. The first potential driver is the cost of living, suggested by the allegation that people migrate out of California because it is too expensive to live in (e.g., Bahnsen, 2021; Beam, 2021). This is in correspondence to the neoclassical economic theory of migration, that migration happens when the move brings net profit, and lower living costs in destination can be a substantial source of net profit. This motivates our hypothesis:

H1: The migration rate from origins with high costs of living to destinations with low costs of living is higher than the reverse.

Following the popular narrative that the California Exodus is a political outcome (Bahnsen, 2021), we hypothesize that political environment could also serve as a driver of population redistribution. Public choice theory and the consumer-voter model consider migration as a means of realizing people's policy preferences (Dye, 1990; Tiebout, 1956). Empirical research on U.S. internal migration has also repeatedly observed Americans "voting with their feet" (Huang and Butts, 2023; Liu et al., 2019; Preuhs, 1999; Tam Cho et al., 2013). The allegation that Californians leaving their liberal state behind are "leftugees" fleeing Democratic governance (Dorsey, 2021) motivates our second hypothesis:

H2: The migration rate from liberal-leaning origins (i.e. those with higher share of supporters for the Democratic Party) towards conservative-leaning destinations is higher than the reverse.

Since population redistribution goes hand in hand with urbanization (Lichter and Brown, 2011; Ravenstein, 1885), it is possible that California Exodus is a reflection of the ex-urbanization process. Henrie and Plane (2008) and Plane et al. (2005) documented the shift of U.S. population from urban areas to rural areas in the 1990s. If this is still happening in 2010s, that might be an underlying mechanism behind California's net migrant loss. We therefore hypothesize that:

H3: The migration rate from urban origins to rural destinations is higher than the reverse.

Last but not the least, racial dynamics play a critical role in American lives, including migration decisions (Crowder et al., 2006, 2012). According to the literature, "White flight" is a frequently observed phenomenon (Boustan et al., 2023; Frey, 1979; Woldoff, 2011), where members of the non-Hispanic White population migrate out of racially-diverse places and settle in White-dominant areas. While White flight is associated with the ex-urbanization process, previous literature has identified racial factors to be a unique and non-negligible contributor to this movement (Frey, 1979; Kruse, 2013). Considering California's diverse racial demographics, White flight could hypothetically contribute to the exodus, and we thus hypothesize that:

H4: The migration rate from origins with low non-Hispanic White concentration to destinations with high non-Hispanic White concentration is higher than the reverse.

These hypotheses embody a combination of conventional wisdom and notions motivated by migration patterns seen elsewhere. But are any of them true - and, more importantly, can they account for the California Exodus? For this, we turn to our empirical analysis.

4.3 Materials and Methods

4.3.1 Data

We model the intercounty migration flow network among all 3,142 U.S. counties. The outcome of interest is the average number of migrants moving between each directed pair of counties each year during 2011-2015, which is calculated and released by the American Community Survey (ACS) administered by U.S. Census Bureau (2019).

The key covariates capture the characteristics of origin and destination in their costs of living, political climates, level of urbanization, and racial compositions, using 2010 Decennial Census and 2006-2010 ACS (U.S. Census Bureau, 2011a,b). The cost of living is measured by the median housing costs in 2006-2010 ACS; the political climate is represented by the percentage of voters that voted for the Democratic candidate (Obama) in the 2008 presidential election, as that was the latest national-level election before the study period. The level of urbanization is indicated by the proportion of rural population of a county, estimated by the 2010 Decennial Census. Lastly, the feature of a county's racial composition is described by its Non-Hispanic White population in the 2010 Census, as this is the most populous racial-ethnic category in the U.S.

The model also considers other covariates that can potentially influence the magnitude of migration flows. The demographic covariates include the (log) population size, log population density (in thousand people per squared kilometers), and age structure (potential support ratio, PSR: ratio of population that are 15-64 years old over population that are 65+ years old), all using 2010 Census Data. The economic covariates include percentage of renters (in contrast to home owners) using 2010 Census, unemployment rates, and percentage of population with higher education attainment, both using 2006-2010 ACS. The geographic covariates include the log distance between origin and destination counties (in kilometers), a dummy variable indicating whether they belong to the same state, and fixed effects for the four major U.S. regions (Northeast, South, Middle West, and West). We also include log migration flow in the previous time period (2006-2010) of the focal migration flow, and the network dependence terms specified in the following section.

4.3.2 Valued TERGMs

We first model the migration patterns using the valued temporal exponential-family random graph models (valued TERGMs, or VTERGMs) (Krivitsky, 2012). The ERGM is a parameteric generative model that impose an exponential family distribution to describe the network structure of interest:

$$\Pr(Y = y|\theta, X) = \frac{h(y)\exp(\theta^T g(y, X))}{\sum_{y' \in \mathcal{Y}} h(y')\exp(\theta^T g(y', X))}, \tag{4.1}$$

where Y is the random variable of network with realization y. $g(\cdot)$ is a vector of sufficient statistics with corresponding parameters θ. The sufficient statistics can be flexibly specified to incorporate both structural covariate effects (e.g., housing price differences and migration flow in the last time point, making it a temporal ERGM or TERGM), and endogenous dependence terms that capture autocorrelations among migration flows. In this chapter, we include two dependence terms, mutuality and waypoint flow, to account for the endogenous mechanisms that contribute to the symmetry at the dyad-pair level and the node level, beyond the specified covariate effects. Mutuality captures the scale of *reciprocated flow* within dyad pairs ($i \to j, j \to i$) by calculating the summation of the minimum edge value across all dyad pairs:

$$g_m(y) = \sum_{(i,j) \in \mathbb{Y}} \min(y_{ij}, y_{ji}). \tag{4.2}$$

The larger the reciprocated flow within a dyad pair, the larger the statistic. For example, if there are 6 migrant exchange between counties i, j, a distribution of {3,3} will have the largest reciprocated flow and the corresponding statistic (3), and a distribution {0,6} will have the smallest (0). Therefore, a positive coefficient will indicate an endogenous pattern of dyad-level reciprocity, and vice versa. The waypoint flow takes a similar formula, but captures the volumetric flow through each node by examining its total inflows and outflows:

$$g_f = \sum_{i \in \mathbb{V}} \min\{ \sum_{j \in \mathbb{V}, j \neq i} y_{ij}, \sum_{k \in \mathbb{V}, k \neq i} y_{ki} \}. \tag{4.3}$$

The larger the volumetric flow moving in and out of a node, the larger the statistic. A positive coefficient will indicate an endogenous pattern of node-level symmetry, and vise versa.

$h(y)$ is a reference measure that determines the probability distribution of the networks when $\theta \to 0$. As a Valued ERGM, since our outcome of interest is the count of migrants between two counties, we specify the shape function as a Poissonian reference measure:

$$h(y) = \prod_{(i,j) \in \mathbb{Y}} (y_{ij}!)^{-1} \tag{4.4}$$

This amounts to the assumption that migration events are indistinguishable within edges. The denominator of the equation 4.1 is the normalizing factor that defined on $\mathcal{Y}$, the set of all possible network configurations based on the same vertex set. This intractable function is the source of computational complexity for ERGMs, as it is a function of both the parameter to be estimated, and the set of possible network structures. This is especially the case for Valued ERGMs, since each dyad now can take not only two values for binary networks, but all natural numbers. The more than three-thousand nodes also increases the computational load of our model. To grapple with this challenge, we employ a parallelizable subsampled Maximum Pseudo-likelihood Estimation procedure for Valued ERGMs (Huang and Butts), which is efficient and shows good estimation quality for high-edge-variance networks such as ours.

4.3.3 Functional Form Specification

There are many possible functional forms for network models even just considering linear formats, since the edge-based models jointly account for the covariates of origin and destination. We thus formulate our key covariate effects based on our theoretical assumptions of their mechanisms that influence migration.

For the cost of living, we include the housing costs of origin and their the difference between destination and origin (destination minus origin). Drawing on the aspiration-ability model of migration (Carling, 2002; Carling and Schewel, 2018), we posit that origin housing costs influences people's financial well-being, which translates into their capacity to migrate; the difference in housing costs influence the utility gain of migrating, altering their aspiration of the migration.

In terms of political, rurality, and racial covariates, we include a dissimilarity measure, implemented as the absolute difference between origin and destination in the corresponding covariate. This follows the operationalization of previous analysis (Huang and Butts, 2023), which reveals a segmental effect in which less migration happens between counties with larger difference in political climates, levels of urbanization, and racial compositions. Since our interest is population redistribution generated from asymmetric migration, we further include two directional effects. The first is the covariate level of the origin, and the second is a sign function (+1 when destination has a higher covariate level than origin, -1 when the reverse, and 0 when equal).

4.3.4 Visualizing Functional Forms

The composite functional forms of each covariate effect pose the question of how to unpack and interpret their joint effects. We develop a visualization protocol that tackles this problem. For each functional form, the protocol calculates the expected edge value under each possible combination of the covariate value of the origin and destination. To make it more comparable across functional forms, we then normalize it by calculating the ratio of this expected value over the expected value that would be obtained if both origin and destination took the average observed value of the covariate.[2] We describe this formula as follows.

In the absence of dependence terms, a Poissonian Valued ERGM is identical to a network regression model with a independent Poisson distributions on each edge (Krivitsky, 2012), where there expected value of the i, j edge is:

$$\mathbb{E}(Y_{ij}) = \exp(\theta^T \Delta_{ij}^{0\rightarrow 1} g(y, X)) \tag{4.5}$$

where Δ denotes the change in the sufficient statistics when the focal edge's value goes from zero to one. If we only focus on one covariate $f_k(X_{ij})$ (whose sufficient statistic in ERGM will be $f_k(X_{ij}) \cdot y_{ij}$), then we have:

$$\mathbb{E}(Y_{ij}) \propto \exp(\theta_k \cdot f_k(X_{ij})) \tag{4.6}$$

so we can express the conditional expected value as a function of origin's and destination's covariate level by calculating the exponentiated product of the functional form and the corresponding coefficient in equation 4.6. We further add a normalizer to center the expected value and make it more comparable across different functional forms. The normalizer is the

---

[2] For political, racial and rurality covariates, we use the population-weighted national mean, treating every county as if it had the same share of Democratic voters, non-Hispanic Whites and rural population as the national percentage. For housing prices, we use the national median, as the functional form takes the logarithm of the prices.

expected edge value when the covariate of the origin and destination is set to the average value (described in the previous footnote) across the vertex set (X_0):

$$\mathbb{E}(Y_{ij}) \propto \exp(\theta_k \cdot [f_k(X_{ij}) - f_k(X_0)]) \tag{4.7}$$

The formula is in essence the ratio between the expected value of a focal edge under a specific origin-destination covariate vector over the expected value where the origin and destination has the covariate value equal to the average value.

When we need to calculate the ratio for composite expected value, we can simply take the product of their ratios for each form. In the Results section, we will display the functional form of both separate effects (e.g. origin housing costs) and composite effects (e.g. origin housing costs plus difference in housing costs).

Note that this is not exactly the same as the conditional expectation ratio in our specified model, since the model contains dependence terms that distort the edge distribution away from a regular Poisson distribution. A rigorous calculation of the exact expectation ratio is, however, computationally prohibitive, as it requires numerical integration of all possible edge values times their probability function for every realization of the covariate vector. Nevertheless, the knockout experiment in the following subsection takes the dependence into control, offering a closer look at the functioning of the Valued ERGM with dependence terms.

4.3.5 Knockout Experiments via Network Simulation

The visualization of functional forms offers structurally “local” insights about how each social force influences migration patterns. Building upon that, we want to quantify how theses social forces contribute to the social phenomenon of interest on a global scale, specifically population redistribution and the California Exodus. We achieve this by leveraging the generative feature of ERGMs to perform *in silico* knockout experiments via network simulation. A *knockout experiment* as employed in a social science context is a model-based thought experiment that examines counterfactual scenarios where certain posited social forces are inoperative, while all other forces are left at their observed levels (Huang and Butts, 2023). The change in outcomes of interest relative to the behavior of the full model is used to probe the impact of the knocked-out mechanism. Here, we implement knockout of mixing effects by simulating migration flows with all counties having their covariates of interest fixed at an identical value average that is specified in the previous footnote (removing differential mixing). Simulating flows obtained under these conditions, we compare California's ranking in net migration loss across all states under the knockout scenarios with the observed models. This allows us to probe the connection between the mechanisms captured by the model and our social phenomenon of interest. For example, if under the hypothetical condition where every U.S. county has the same housing cost, California's relative net migration loss is not as severe as the observed situation, it would suggest that housing-cost effects on migration could be a contributor to the California Exodus; by turns, if eliminating housing disparities has no impact on asymmetric migration, we can rule it out as a driver of migration loss.

To assist the interpretation of the quantitative results from knockout experiments, we include positive and negative controls in simulation, alongside knockouts of our key covariates of interest: political, racial, rurality, and housing attributes. Originating in the experimental sciences, positive and negative controls are experimental conditions that researchers expect to produce positive and null results, respectively; the controls validate the experimental procedures, serving as the benchmark for other regular experimental settings. In an *in silico* setting, controls remain important to verify that the model is sensitive to manipulations that should have an impact on the outcome of interest (and, by turns, that it is not overly sensitive to manipulations that should not have an impact). Here, we knock out distance effects as a negative control, treating all dyads as having a common log distance set at national mean. We expect the knockout of non-directional distance effects to not alter the rankings of net migration loss across the country, and the difference between this case and the full model can be considered as a combination of numerical noises and some second-order impacts (since we include complex network dependence terms). The removal of population effects by equally distributing population across all counties serves as a positive control case, as we expect the removal of population effect to have a large impact on the population redistribution pattern. The purpose of these two controls is not substantive interpretation of the fundamental distance and population effects, as the counterfactual scenario is arguably radical and unrealistic, but rather, to provide insights into the question of "how small is small" and "how big is big" in terms of altering migration ranking.

4.4 Results

4.4.1 General Patterns of Population Redistribution

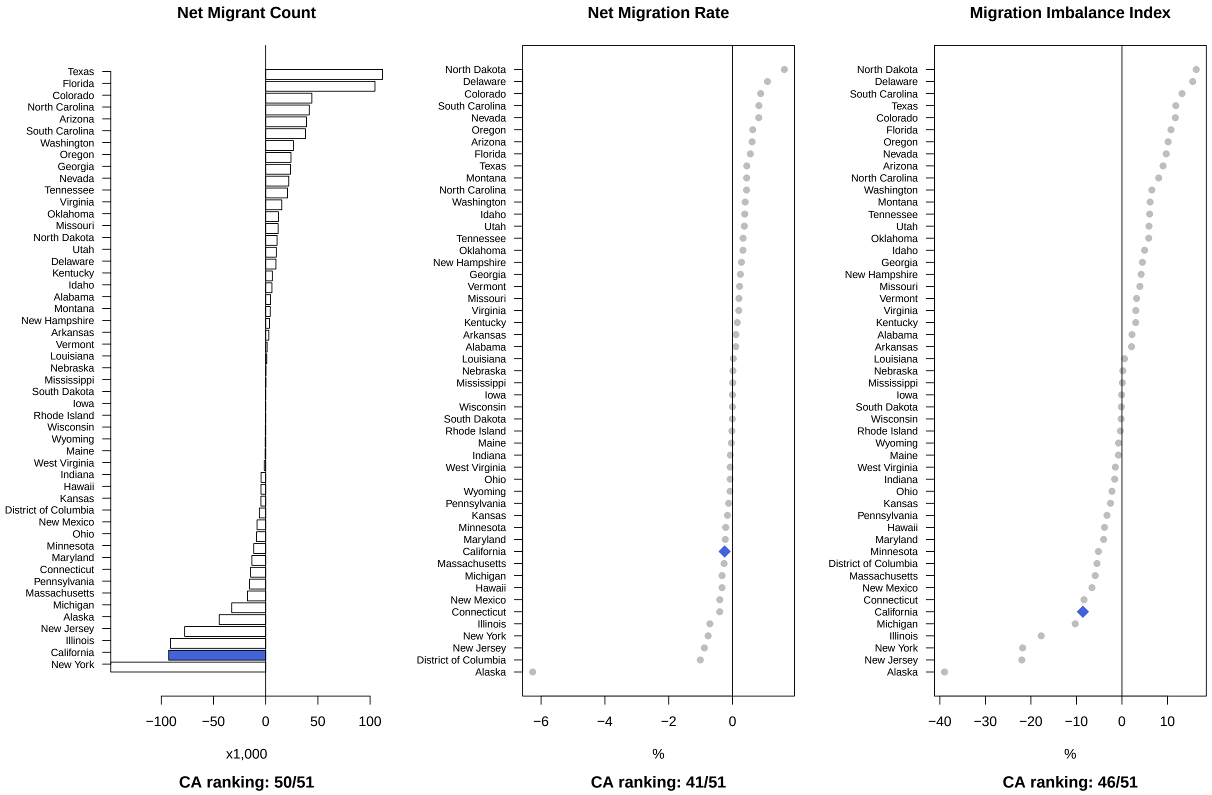
Table 4.1: Annual Population Change in the United States, 2011-2015

	Count	Crude Rate (%)
Population	308,739,316	
Natural Change		
Births	3,961,037	1.28
Deaths	2,598,956	0.84
Natural Increase	1,362,081	0.44
International Migration		
Immigration	1,841,695	0.60
Intercounty Migration		
Total migrants	17,176,675	5.56
Node-level asymmetry	1,523,550	0.49
Dyad-level asymmetry	3,844,434	1.25

To offer a broad view of population change in the study period, Table 4.1 shows the annual population changes from different demographic processes and their crude rates (normalized by the total population size).[3] Compared to natural change and international migration,

---

[3] The population size comes from 2010 Census, the natural change data comes from U.S. Center for



Net Migrant Count

Net Migration Rate

Migration Imbalance Index

CA ranking: 50/51

CA ranking: 41/51

CA ranking: 46/51

Figure 4.1: Net migrant count, net migration rate, and migration imbalance index by state

intercounty migration in the U.S. is a more substantial demographic process with a larger share of population involved. When it comes to population change, the asymmetric internal migration is similar to the scale of immigration and natural increase, all of which have a modest share of population, whcih is around 0.5% to 1%. This confirms that as a developed country, the U.S. has a relatively modest population change in the 2010s (Rees et al., 2017).

Figure 4.1 examines the phenomenon of the California Exodus by comparing the net migrant loss of California (shaded in blue) across three metrics against other U.S. states and the District of Columbia (DC). The left panel displays the net migrant count, which is the total in-migrants minus the total out-migrants. It shows that California has a large net migrant

---

Diease Control and Prevention, and the international and internal migration comes from ACS 2011-2015. The natural increase is the number of births minus the number of deaths. The dyad-level asymmetry is the sum of absolute difference across all dyad pairs divided by two: $A_d = \frac{\sum_{i,j} |Y_{ij} - Y_{ji}|}{2}$, and the node-level asymmetry is the sum of absolute difference across all nodes in their inflows and outflows divided by two: $A_n = \frac{\sum_i | \sum_i Y_{ij} - \sum_i Y_{ji}|}{2}$.

loss, only second to New York among the 51 states and DC.

Yet, considering the fact that California is the most populous state (roughly 25% more than the second populous state, Texas, in 2010), the middle panel calculates the net migration rate, which is the net migrant count divided by the state's population. The normalized metric observes California to have a less extreme net migration loss. While it still ranks at the lower end of the list, it is not very different from the majority of the U.S. states, which are within the range of -1% to 1%. In other words, the large net outflows of migrants from California can be partly explained by its largest population size.

Although the middle panel may suggest that there is nothing to be explained - the California Exodus is simply a size effect - examining the *relative asymmetry* of migration to and from California gives a richer picture. The right panel calculates the migration imbalance index (MII) of each state, which is the net migrant count divided by the sum of in-migrants and out-migrants.[4] The measurement indicates the proportion of related migrant flows that are inflows of a focal place, capturing the level of imbalance between inflows and outflows of migrants. The right panel reveals that migration imbalance generally has larger variation across states than the net migration rate, as the former focuses on a smaller population, i.e. the migrant population. California has relatively lower ranking in migration imbalance than net migration rate, and its value is farther away from other U.S. states, suggesting a noticeable imbalance in its in/out-migration flows.

In summary, Figure 4.1 reveals that California is indeed experiencing net migration loss, although the severity relative to other parts of the country vary by the metric we read. Moreover, despite the popularity of the California Exodus narrative, California is actually not the place with the most net migration loss: the New York state has stronger net loss

---

[4] MII coincides with the migration efficiency/effectiveness index in some migration literature (Bell et al., 2002; Shryock et al., 1973). It is also directly related to the external-internal (E-I) Index in social network analysis (Krackhardt and Stern, 1988), although the latter focus on external flows, so MII is equal to one minus the E-I index.

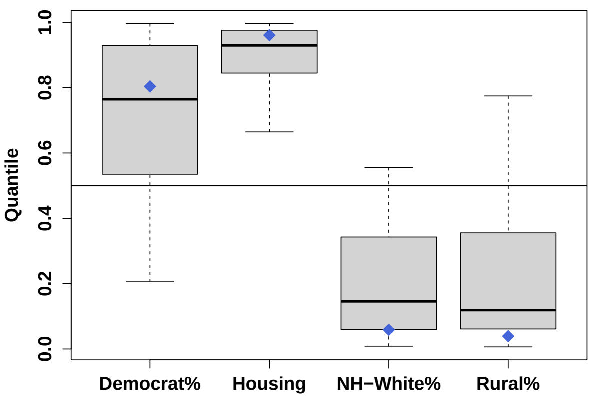


Figure 4.2: Quantiles of California's attributes
Note: Boxes represents California counties, and blue diamonds represents the whole California state (weighted by county's population size).

than California across all metrics, and the net migration rate and migration imbalance of Alaska is substantially lower than the rest of the states. These other cases poses important empirical questions that future research should consider.

Lastly, as we consider the possible contributor of California's outstanding net migration loss, we examine California's attributes in Figure 4.2. The boxplots shows the quantiles of California counties in those attributes across all U.S. countires, and the blue diamonds indicate the quantiles of California across the 51 states and DC. Compared to other parts of the country, California is indeed a place with stronger left-leaning political environments, expensive housing, larger racial and ethnic minority population share, and higher levels of urbanization. These dimensions are characteristics where California stands out, and therefore has the potential of explaining its migration patterns.

Table 4.2: Valued TERGM for Intercounty Migration Flows, 2011-2015

	Estimate	Std Err
Political Covariates		
Dissimilarity P(Democrat)	-.257***	.007
Origin P(Democrat)	.024**	.009
To higher P(Democrat)	-.008***	.001
Racial Covariates		
Dissimilarity P(NH-White)	-.172***	.006
Origin P(NH-White)	-.044***	.007
To higher P(NH-White)	.011***	.001
Rurality Covariates		
Dissimilarity P(rural)	-.457***	.004
Origin P(rural)	.330***	.006
To higher P(rural)	.018***	.001
Housing Covariates		
Origin log(costs)	-.283***	.005
Difference log(costs)	-.148***	.004
Control Covariates	(included)	

Note: **$p < 0.01$; ***$p < 0.001$ (two-tailed tests).

4.4.2 Functional Forms of Migration Driving Forces

Estimated Effects

To explain the underlying patterns of intercounty migration, we estimate a VERGM for the migration flow network, with the results of the key covariates of interest listed in Table 4.2. The model suggests that, on average, less migration happens between counties with larger differences in their political climates, rurality, and racial compositions as reflected by the percentage of the non-Hispanic White population. In terms of directional effects, the model predicts larger migration flows from counties with higher Democratic Party voter share, and towards counties where the Democratic party voter share is lower. The directionality of the political effects is largely in correspondence to the "lefugee" Hypothesis 2 that population are generally leaving from Democratic-party-leaning areas towards Republican-party-leaning areas. The racial effects also run in the direction predicted by the "White flight" Hypothesis 4. Holding other factors constant, counties with smaller proportions of non-Hispanic White population send out more migrants, and larger migration flows exist along the way that lead to a county with a higher share of non-Hispanic White population.

When it comes to rurality, the model is consistent with the ex-urbanization Hypothesis 3 that migration flows are larger when they are moving towards counties with a higher share of rural population than the origin. Yet, the model also shows that counties with higher rurality on average send more migrants out than those with lower rurality. In other words, more migration flows are moving towards more rural regions, but more of them come from a rural county. The housing effects also offer mixed evidence in light of the neoclassical-economic Hypothesis 1. Although migration flows are larger where moving brings greater declines in housing costs from origin to destination, counties with lower housing costs also observe larger out-migration flows. This means that migration typically happens from places with inexpensive housing to places with even less expensive housing.

The model also controls for a series of other covariate effects and endogenous dependence structure, reported in Table 4.A in the Appendix. The positive mutuality and the negative waypoint flow patterns suggest that, holding other covariate effects constant, the observed migration flow network is more reciprocal at the dyad-pair level and less symmetric at the node level than a random network configuration. This implies the existence of endogenous network patterns discussed in prior literature (Leal, 2021; Huang and Butts, 2023). For example, the practice of return migration could promote dyad-level reciprocity, and the signaling effects of county attractiveness can lead to endogenous node-level asymmetry (large migration inflows signaling the popularity of this county, retaining potential migrants from leaving, resulting in an imbalanced in&out-flow of the county).

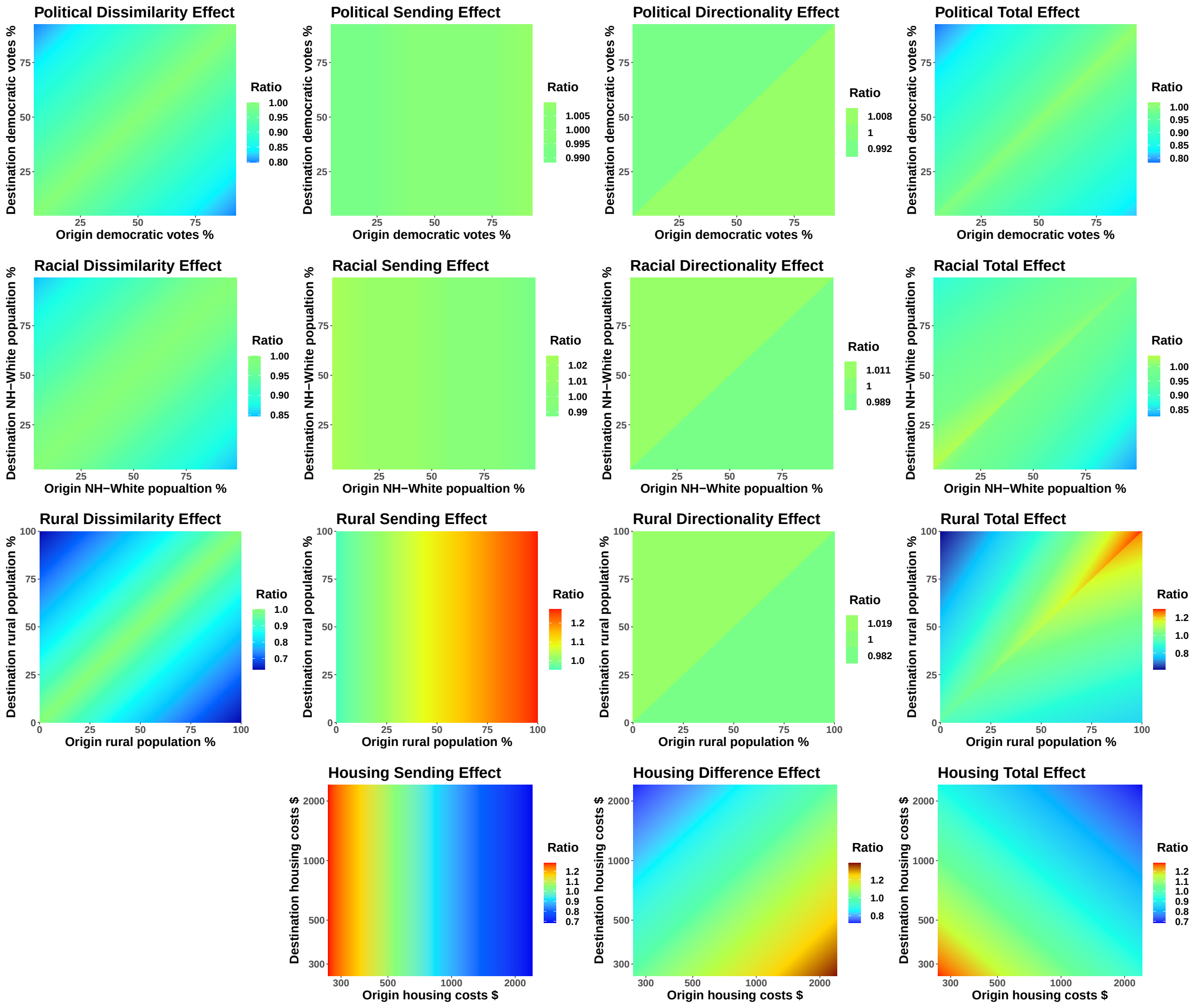


Figure 4.3: Function forms for political, rural, racial and housing effects

Visualizing Functional Forms

For a typical research paper using parametric models, the results section usually stops at the previous subsection, after summarizing whether the directionality of the key effects confirms or refutes the hypothesis. While it is informative to use parametric models as tools for hypothesis testing by evaluating their qualitative behavior, there are more insights one could gain from further the examination of the models.

First of all, besides the signs of the coefficients and their corresponding p-values, their magnitudes also carry critical information about the scale of the effects of interest. Taking the political covariates in Table 4.2 as an example, the coefficient of origin effects and binary directional effects look an order of magnitude smaller than that of the dissimilarity effect. However, it is difficult to directly interpret the parameter magnitudes, which is subject to the scaling of the covariate distribution.

The second question is about how to interpret holistically the effects of interest, as the different effects (origin, difference, dissimilarity) are interdependent, and holding other factors constant to interpret each single functional form can be unrealistic. This could be a critical question as sometimes different effects offer mixed evidence about substantive hypotheses, such as the rurality and the housing effects in our model. It is of substantive interest to understand how these different effects jointly shape the migration pattern.

To quantify the magnitude of the modeled effects and more concretely understand the separate and joint roles of the functional forms, we visualize the (normalized) predicted migration flow size as a function of origin's and destination's covariate values, displayed in Figure 4.3. Each row presents one chunk of covariate effects, and each column presents a type of functional form, where the higher value in the heatmap indicates the model predicts the migration flow to be higher under these origin-destination covariate values.

The first row of Figure 4.3 shows that the directional functional forms (sending and directionality effects) produce very little alternation of the expected migration flow, compared to the undirectional functional forms (dissimilarity effects). The middle two panels show a tiny gradient in its coloring, and the total effects largely resemble the dissimilarity effect, suggesting that the sending and directionality effects make little contribution to the overall effect of political climate. Similarly, in the second row, directional effects of racial covariates also appear negligible, and the undirectional dissimilarity effect dominates the total effect of racial composition. These visualizations tell us that while the directional effects of political and racial covariates run in the direction that correspond to the hypotheses, their effect sizes are small compared to the nondirectional dissimilarity effects.

In the third row of Figure 4.3, although the directionality effect of rurality still resembles those of the previous political and racial covariates, bringing small variation in the expected migration scale, the rural sending effect is strong, and alters the rural total effect to be asymmetric. The bottom row shows that while the sending and difference effects predict substantial variation of expected value across different housing values, their combination offsets each other in the bottom right panel; the gradient of the total effect largely evolves along the $y = x$ line, meaning that swapping the housing costs of origin and destination does not leads to major change in the expected migrant counts. This means that the total housing effect is largely symmetric.

Visualizing Functional Forms: The San Francisco County Case

To further aid our interpretation of the total effects, Figure 4.4 examines the case of San Francisco (SF) county, California, and evaluates its expected migration flows towards and from other counties based on their corresponding covariate value. The first column is a replication of the last column in the previous figure, but adds reference lines that indicate the covariate level of SF county. The middle column extracts from these two reference lines and plot the expected number of immigrants to (brown solid lines) and emigrants from (grey dotted lines) SF county as a function of the origin/destination county's covariate level. The upper right panel of each row summarizes the middle column by getting the difference of immigrant ratio and the emigrant ratio, where a positive ratio difference (shaded in solid brown lines) suggests an expected net migration gain for SF county, while a negative ratio difference (shaded in dotted grey lines) suggests an expected net migration loss for SF county. The bottom right panel of each row plots the histogram of U.S. population about the covariate level of their residing counties. The juxtaposition of the last two plots reflects whether the country's population gravitate towards counties that SF county has net

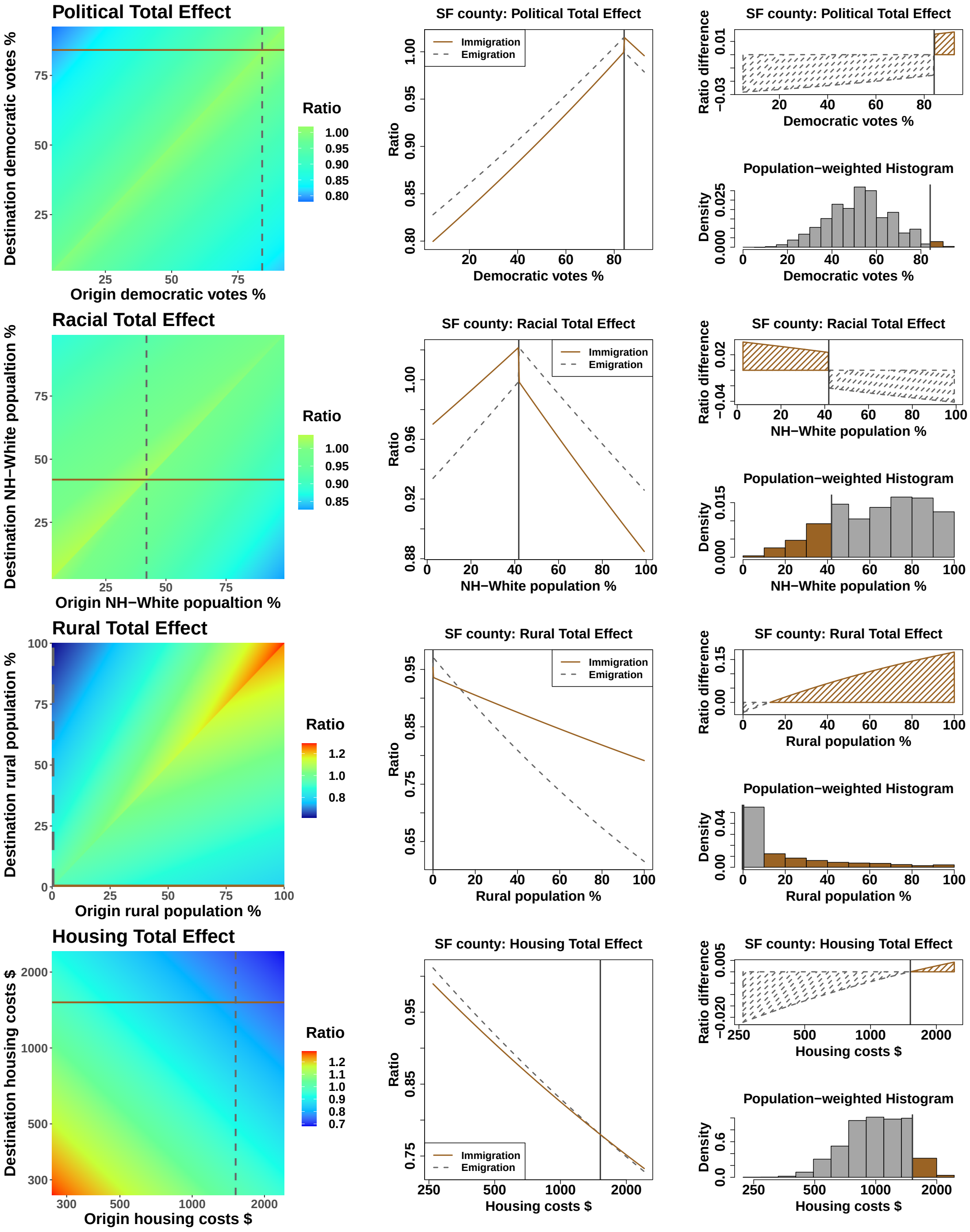


Figure 4.4: Function forms for migration effects involving San Francisco county.

Note: (left) Dyadic effects, with vertical and horizontal lines showing SF attributes. (center) Net immigration (solid lines) and emigration (dotted lines) effects for SF, given origin/destination county attributes; vertical line shows SF position. (right) Areas between curves (net immigration) from the center plot by origin/destination county attributes; histograms show population-weighted distributions of U.S. counties, with brown columns indicating population in net SF-immigration counties.

migration gain from (shaded in brown), or counties that SF county has net migration loss towards (shaded in grey), offering a first-order approximation to whether the social effects promote or suppress population loss from a county like San Francisco.

Focusing on the right column of Figure 4.4, we observe that SF county receives net migration gains from counties with more Democratic-party voter share, which comprise a small share of U.S. population. By turns, it loses migrants to counties with less Democratic-party voter share, which comprise a large share of U.S. population. Similarly, in the second row, SF county receives net migration gains from counties with less non-Hispanic White population share, which comprise a small share of U.S. population. The functional form of rurality for SF county is a bit more complicated, as the county takes the extreme value of 0% rural population. The county is expected to have no net migration exchange with other counties that have 0% rural population, which consist 7% of the total U.S. population. SF county is expected to lose population to counties with rural population larger than zero but smaller than 13%, which includes about 51% of the total U.S. population. In other words, on average, there are slightly more persons residing in counties that SF county has net migration loss towards. However, once the county deviates from the extreme case of the fully urbanized, the trend reverses, with more of the U.S. population residing in places from which the focal county has net migration gain. Lastly, the bottom right panel shows that the majority of the U.S. population resides in counties with cheaper housing than SF county, areas to which SF would be expected (*ceteris paribus*) to lose population. Overall, for the SF county case, across all covariates, the model predicts an overall net migration loss from SF county; this is not because all factors *unilaterally* favor emigration from SF, but rather because in each case SF's attributes favor immigration from a relatively small number of counties (with relatively low total population) relative to those to which they favor emigration.

Table 4.3: California's Average Simulated Ranking with and without Knockouts

	Net Count		Net Rate		Imbalance Index	
	Ranking	Change	Ranking	Change	Ranking	Change
Full Model	50.00		42.08		45.55	
Remove Distance Effect	50.00	0.00	41.92	-0.16	45.35	-0.20
Remove Population Effects	48.75	-1.25	28.30	-13.78	46.63	+1.08
Remove Political Effects	50.00	0.00	42.00	-0.08	45.16	-0.39
Remove Housing Effects	50.00	0.00	41.92	-0.16	44.95	-0.60
Remove Racial Effects	50.00	0.00	39.36	-2.72	42.49	-3.06
Remove Rurality Effects	50.00	0.00	43.00	+0.92	47.49	+1.94

4.4.3 Knockout Experiments for the California Exodus

The visualization of covariate effects offers us some quantitative insights about how different social forces operate across different origin/destination pairs. However, our examination of the SF case underscores the intuition that the way in which such forces play out depends upon the global distribution of population (and covariates), which is challenging to infer from direct inspection. For instance, the high level of urbanization in SF county makes it an interesting but special case, and it becomes difficult to visualize every possible rurality level that California counties take and integrate them to offer a holistic evaluation of the rurality effect on the California Exodus. Building on these exploratory insights, this section aims to explicitly examine the connection between migration patterns incorporated into the model with specific social outcomes of interest, such as the California Exodus.

We achieve this by performing *in silico* knockout experiments, with results displayed in Table 4.3. The first column suggests that California's ranking in net migrant count stays constant throughout all the knockout scenarios except the positive control that knocks out population, contributing to a 1.25 position improvement its ranking (smaller ranking means less net migrant loss). Notice that only knocking out population effects in the positive control alters California's average ranking in net migrant count, and that in the second column, the net migration rates under normalized state population lead to fluctuations of California's average rankings under all knockout scenarios. This suggests that California's status as the largest U.S. state is a major explanation for its substantial net emigration in absolute terms.

In the second column of Table 4.3, the removal of political and housing effects improves California's ranking in net migration rate at a scale smaller than or roughly equal to the negative control of removing distance effects. Although political and housing effects seem to operate in a direction that contributes to California exodus as hypothesized, their influence on net migration rate is substantively negligible. Knocking out racial effects and rural effects improves and worsens California's relative net migration rate, respectively, indicating that racial effects contribute to California Exodus (from a migration rate angle), while rural effects actually buffer California from even larger population loss. These two changes are larger in their scale than the negative control of distance effects, but not comparable to the positive control of population effects, suggesting their impacts to be moderate.

The last column in Table 4.3 shows California's ranking of migration imbalance. As with the case of net migration rate, removing political, housing, and racial effects reduces California's relative migration imbalance, while removing rurality effects worsens it. Quantitatively speaking, the impact of knockouts of political and housing effects are again similar to that of the negative control of distance effect, while the removal of racial and rural effects bring a ranking change even larger than that from the positive control case of population effects. The small alteration from the positive case is understandable, as the origin and destination effects of population are not hugely different in our model (as well as in many other gravity models, Boyle et al. (2014)); while changing the total size of migrant population (symmetrically) can alter state rankings of net migration rate given a constant total population denominator, for migration imbalance that solely focuses on the migrant population, this is no longer the case. The fact that none of the knockouts alters California's relative migration imbalance in a sizable way suggests that California's migration imbalance does not result from one single social effect.

4.5 Discussion

Leveraging a large-scale valued network model, this chapter studies population redistribution patterns in the United States, and in particular the heatedly discussed case of the "California Exodus." Our analyses show that California indeed experienced net migration loss in the 2010s, although its scale varies depending on the metrics one examines; the exodus is substantial in absolute terms but relatively small in its crude rate (count per capita), while still being fairly considerable in its imbalance between in-migration and out-migration flows. Valued TERGM analysis reveals the direction of the political, rural, racial, and housing effects on population redistribution, which largely work in directions that would contribute to net migration loss for highly populous counties like San Francisco. Knockout experiments further show that racial effects contribute to the California Exodus, rurality effects work *against* the California Exodus, and while political and housing effects contribute to the California Exodus, their effects are largely negligible. The scale of these effects on the California Exodus varies by the migration metric used, but none of the knockout scenarios (except a positive control case for population distribution) alter California's ranking in net migration loss in a substantial way. This suggests that the California Exodus is not governed by one single social effect, but is a joint outcome of complex systemic patterns.

Methodologically, this chapter offers a roadmap that aids interpretation of composite functional forms in parametric relational models via visualization. It also demonstrates the insights generative models such as ERGMs could offer by designing simulation experiments for relevant counterfactual questions. In our view, this provides a reminder that network models are not merely statistical hypothesis-testing tools, but flexible and powerful generative devices that can reveal emergent effects of multiple mechanisms on outcomes of interest in complex social systems.

In closing, we note that while statistical network models have seen great advances over the past 20 years, important challenges remain. Among these is the problem of accounting for measurement error (a persistent challenge for the field since the famous call-to-arms of Bernard et al. (1984)). As with the vast majority of work in both social network analysis and demography, this chapter considers the data as a fixed input without accounting for measurement error. However, even Census data is imperfectly measured, a concern that becomes greater when considering the $\mathcal{O}(3000^2)$ migration rates that must be estimated to measure the U.S. county-level migration system. Assessing the nature and consequences of measurement error in migration networks remains an open problem, as does the estimation of count-valued ERGMs in the presence of measurement error. These would seem to be important directions for further work.

Likewise, in defining a network, one's choice of nodes and edges imposes a certain level of granularity on one's representation, which in turn impacts what effects it can distinguish (Butts, 2009). Here, we examine the network of migration flows among U.S. counties, which could itself be seen as an aggregation of an ensemble of migrant flow networks for smaller subsets of the U.S. population; although we can hypothesize how these subflows contribute to the aggregate flow network, we are limited in our ability to disaggregate them here. For example, we do not have information about whether and to what extent the larger migration flow from low-White-concentration counties to higher-White-concentration counties is driven by movement of the non-Hispanic White population, versus members of minority populations following on the heels of earlier migration by non-Hispanic Whites (an effect seen in some past research, e.g. Woldoff (2011)). Distinguishing the migration patterns of different population groups within a joint model imposes significant challenges both from a data availability/accuracy and computational standpoint, but could provide further insights if feasible.

Last but not least, we note that there exist other states whose population redistribution patterns are stronger than California, such as New York State and Alaska, despite receiving less public attention. The impacts of the pandemic on internal migration, over both the short term and the long term (e.g., potential enhancement of ex-urban migration), are also critical research topics. We encourage future research to examine these cases to offer a more comprehensive understanding regarding the evolution of the U.S. migration system and its implications for American society.

4.6 Appendix

Table 4.A: Valued TERGM for Intercounty Migration Flows, 2011-2015 (Full Model)

	Estimate	Std Err
Key Covariates in Table 2	(included)	
Dependence Structures		
Mutuality	.047***	.002
Waypoint flow	-.013***	.001
Log(past migrant flow)	.300***	<.001
Demographic Covariates		
Origin log(population size)	.353***	.002
Origin log(population size)	.374***	.002
Destination log(population density)	-.081***	.001
Origin log(population density)	-.055***	.001
Destination PSR	.017***	.001
Origin PSR	.017***	.001
Destination log(immigrant inflow)	.057***	.001
Origin log(immigrant inflow)	.043***	.001
Economic Covariates		
Destination P(higher education)	.386***	.012
Origin P(higher education)	.314***	.013
Destination P(renter)	.404***	.012
Origin P(renter)	.421***	.012
Difference P(unemployment)	-1.229***	.041
Origin P(unemployment)	-2.948***	.052
Geographical Covariates		
Log(distance)	-.569***	.001
Same state	.500***	.002
Northeast	(reference group)	
Destination South	.257***	.003
Origin South	.053***	.003
Destination West	.384***	.004
Origin West	.225***	.004
Dstimation Midwest	.202***	.003
Origin Midwest	.096***	.003
Baseline		
sum	-1.421***	.042
nonzero	-13.965***	.028

Note: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$ (two-tailed tests).

Chapter 5

Conclusion

This dissertation examines the social forces that drive and inhibit migration. Using exponential-family random graph models (ERGMs) for valued networks, we model the structure of inter-U.S.-county migration flows from a systemic and relational perspective. The dissertation contributes to the fields of social networks, migration, spatial demography, and methodology in the following ways:

- Chapter 2 proposes and implements a fast and accurate parameter estimation procedure for count-valued ERGMs, removing the computational hurdle to model large-scale high-edge-variance network systems. The comparative simulation experiments further show how the network size and the variance of edge values influence the performance of the available computational methods, providing a guideline for choosing and tuning those methods for different use cases.
- Chapter 3 identifies and quantifies how segmentation across geography in political environment, racial composition, and urbanization can contribute to the population immobility in the United States. It advances the scholarship of immobility with theorization and empirical evidence of the connections between segregation, polarization,

and (im)mobility.

- Chapter 4 develops an analytical framework to study migration-induced population redistribution and its underlying mechanisms. It offers empirical evidence to a popular public discourse about the scale and the drivers of migration-induced population redistribution between California and other U.S. states.

- The dissertation demonstrates the power of using network generative models such as ERGMs to study social systems. When studying flow systems consisting of exchange of people, information or goods among places, organizations, occupations or other social categories, network models can describe the interactive effects of origin and destination on the flows. The models can also reveal the endogenous mechanisms in the system by describing the structure of interdependence among flows. Furthermore, we show that generative models are more than hypothesis testing tools and can be used to probe mechanisms via *in silico* experiments, exemplified by the knockout experiments.[1]

The dissertation has the following limitations:

- The analyses are based on one dataset about inter-U.S.-county migration flows during 2011-2015, so results might not be generalizable to other settings. Simulation experiments in Chapter 2 do not guarantee that methods evaluated will have the same performance when applied to data with different structures or properties. Chapter 3 reveals segmentation as a contributor to U.S. population immobility during 2011-2015, but the lack of longitudinal data prevents us from addressing whether and how much it can explain the decline of migration rates in the past decades. Similarly, empirical results in Chapter 4 might not be identical to migration and population redistribution patterns in other times, such as during and after the COVID-19 pandemic in 2020s.

---

[1] We recommend future users of *in silico* experiments read our notes in Appendix (3.7.3).

- The analyses use aggregate-level data of population flows without information about individual behaviors. While these analyses can directly speak to the macrosociological questions of interest about population immobility and population redistribution across geography, they do not capture the micro-level processes behind. Qualitative and quantitative data at individual, household, and/or community levels is needed to answer questions such as how prospective migrants weigh different aspects of potential destinations, how family members negotiate about moving, and how migration aspiration translates or fails to translate to migration behavior.

- Although we set up the Valued ERGMs at their state of the art and the limit of computation feasibility, it may not be the optimal way of modeling the structure of population flow networks. Since the microscopic dynamic process that can give rise to the dependence structure specified in Valued ERGMs is still not well understood, the functional forms used to describe reciprocity and waypoint flows might not be the most effective and sensible ones. The computational hurdle also deters the consideration of higher-order dependence structure such as transitivity of population flows.

- The models do not fully account for uncertainty. While migration flow data from the American Community Survey is an estimate, the models treat the data as fixed input without measurement errors, due to the lack of information about the covariance structure of those errors, and a lack of methods that deal with measurement errors in Valued ERGMs. Moreover, as Duxbury (2023) notes, our knockout experiments only use point estimation of coefficients, without considering uncertainties from coefficient estimations, which demands another level of computational capacity for this large valued network system.

These limitations suggest the following directions for future research:

- Study population flow systems in other time and space, and in longitudinal analysis when data allows. For U.S. internal migration, fruitful questions include: 1) whether and how the decline in migration rates over the past decades is related to the growing political polarization across geography and other spatial segmentation patterns; 2) how the long-term decline in migration rates impact various aspects of social lives; 3) what explains the substantial net migration loss of Alaska to other U.S. states during 2011-2015 and whether similar trends happen in other times; and 4) how the internal migration and population redistribution patterns have changed since the COVID-19 pandemic.

- Based on the population-level migration patterns revealed in this dissertation, collect and analyze micro-level data to triangulate the findings, such as studying how people in the U.S. understand and evaluate political climate, rural-urban divisions, racial dynamics, and housing costs when deciding about and practicing migration.

- Advance the theory, method, and computation of valued network models such as Valued ERGMs. This includes: 1) deriving stochastic processes that can give rise to dependence structures in valued networks and their corresponding specifications in ERGMs and/or other models, 2) developing methods that account for measurement errors for valued networks in both data collection and data analysis stages, 3) implementing even more time-efficient parameter estimation procedure. An example of the last point is to improve uncertainty measurement of pseudo-likelihood-based estimation for networks with small edge variance to broaden its applicability.

Bibliography

Abramski, K., Katenka, N., and Hutchison, M. (2020). A Network-Based Analysis of International Refugee Migration Patterns Using GERGMs. In *Complex Networks and Their Applications VIII*, Studies in Computational Intelligence, pages 387–400, Cham. Springer International Publishing.

Aicher, C., Jacobs, A. Z., and Clauset, A. (2015). Learning latent block structure in weighted networks. *Journal of Complex Networks*, 3(2):221–248.

Akbari, H. (2021). Exploratory social-spatial network analysis of global migration structure. *Social Networks*, 64:181–193.

Aksoy, O. and Yıldırım, S. (2024). A model of dynamic flows: Explaining turkey's inter-provincial migration. *Sociological Methodology*, 54(1):52–91.

Almquist, Z. W. and Butts, C. T. (2015). Predicting Regional Self-Identification from Spatial Network Models. *Geographical Analysis*, 47(1):50–72.

Altman, D. G. and Royston, P. (2006). The cost of dichotomising continuous variables. *BMJ*, 332(7549):1080.

An, W. (2016). Fitting ERGMs on big networks. *Social Science Research*, 59:107–119.

Anderson, C. J., Wasserman, S., and Crouch, B. (1999). A p^* primer: logit models for social networks. *Social Networks*, 21(1):37–66.

Anderson, J. E. (2011). The Gravity Model. *Annual Review of Economics*, 3(1):133–160.

Asuncion, A. U., Liu, Q., Ihler, A. T., and Smyth, P. (2010). Particle Filtered MCMC-MLE with Connections to Contrastive Divergence. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*.

Azose, J. J. and Raftery, A. E. (2015). Bayesian Probabilistic Projection of International Migration. *Demography*, 52(5):1627–1650.

Azose, J. J. and Raftery, A. E. (2018). Estimating large correlation matrices for international migration. *The Annals of Applied Statistics*, 12(2):940–970.

Badger, E., Bui, Q., and Katz, J. (2018). The Suburbs Are Changing. But Not in All the Ways Liberals Hope. *The New York Times*.

Bahnsen, D. L. (2021). The Great California Exodus. *National Review*, LXXIII(7).

Bakewell, O. (2010). Some Reflections on Structure and Agency in Migration Theory. *Journal of Ethnic and Migration Studies*, 36(10):1689–1708.

Bakewell, O. (2014). Relaunching migration systems. *Migration Studies*, 2(3):300–318.

Bakewell, O., De Haas, H., and Kubal, A. (2012). Migration Systems, Pioneer Migrants and the Role of Agency. *Journal of Critical Realism*, 11(4):413–437.

Bakewell, O., Engbersen, G., Fonseca, M. L., and Horst, C. (2016a). *Beyond Networks: Feedback in International Migration*. Springer.

Bakewell, O., Kubal, A., and Pereira, S. (2016b). Introduction: Feedback in Migration Processes. In Bakewell, O., Engbersen, G., Fonseca, M. L., and Horst, C., editors, *Beyond Networks: Feedback in International Migration*, Migration, Diasporas and Citizenship, pages 1–17. Palgrave Macmillan UK, London.

Baldassarri, D. and Gelman, A. (2008). Partisans without Constraint: Political Polarization and Trends in American Public Opinion. *American Journal of Sociology*, 114(2):408–446.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press.

Bauman, Z. (2000). *Liquid Modernity*. John Wiley & Sons.

Beam, A. (2021). California's growth rate at record low as more people leave. *AP NEWS*.

Bell, M., Blake, M., Boyle, P., Duke-Williams, O., Rees, P., Stillwell, J., and Hugo, G. (2002). Cross-national comparison of internal migration: issues and measures. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 165(3):435–464.

Bernard, H. R., Killworth, P., Kronenfeld, D., and Sailer, L. (1984). The Problem of Informant Accuracy: The Validity of Retrospective Data. *Annual Review of Anthropology*, 13(1):495–517.

Bernard, H. R., Killworth, P. D., and Sailer, L. (1979). Informant accuracy in social network data IV: A comparison of clique-level structure in behavioral and cognitive network data. *Social Networks*, 2(3):191–218. Publisher: Elsevier.

Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225.

Bhamidi, S., Bresler, G., and Sly, A. (2008). Mixing time of exponential random graphs. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 803–812. IEEE.

Biagi, B., Faggian, A., and McCann, P. (2011). Long and Short Distance Migration in Italy: The Role of Economic, Social and Environmental Characteristics. *Spatial Economic Analysis*, 6(1):111–131. Publisher: Routledge _eprint: https://doi.org/10.1080/17421772.2010.540035.

Bishop, B. and Cushing, R. G. (2009). *The Big Sort: Why the Clustering of Like-minded America is Tearing Us Apart*. Houghton Mifflin Harcourt.

Block, P., Stadtfeld, C., and Robins, G. (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.

Borjas, G. J. (2006). Native Internal Migration and the Labor Market Impact of Immigration. *Journal of Human Resources*, 41(2):221–258.

Boustan, L., Cai, C., and Tseng, T. (2023). JUE Insight: White flight from Asian immigration: Evidence from California Public Schools. *Journal of Urban Economics*, page 103541.

Boyle, P., Keith H., H., Vaughan, R., and Vaughan, R. (2014). *Exploring Contemporary Migration*. Routledge, Abingdon, United Kingdom.

Briggs, X. d. S., Popkin, S. J., and Goering, J. (2010). *Moving to Opportunity: The Story of an American Experiment to Fight Ghetto Poverty*. Oxford University Press.

Brown, J. R. and Enos, R. D. (2021). The measurement of partisan sorting for 180 million voters. *Nature Human Behaviour*, 5(8):998–1008.

Brown, S. K. and Bean, F. D. (2016). Conceptualizing Migration: From Internal/International to Kinds of Membership. In White, M. J., editor, *International Handbook of Migration and Population Distribution*, International Handbooks of Population, pages 91–106. Springer Netherlands, Dordrecht.

Burris, V. (2004). The Academic Caste System: Prestige Hierarchies in PhD Exchange Networks. *American Sociological Review*, 69(2):239–264.

Butts, C. (2023). In silico experiments with relevent. Retrieved May 10, 2024 (`https://statnet.org/workshop-relevent-experiments/in_silico_experiments_tutorial.html`).

Butts, C. T. (2003). Predictability of large-scale spatially embedded networks. In *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers*, pages 313–323. National Academies Press, Washington, D.C.

Butts, C. T. (2007). Models for Generalized Location Systems. *Sociological Methodology*, 37(1):283–348.

Butts, C. T. (2008). Social Network Analysis with sna. *Journal of Statistical Software*, 24(6).

Butts, C. T. (2009). Revisiting the Foundations of Network Analysis. *Science*, 325(5939):414–416. Publisher: American Association for the Advancement of Science.

Butts, C. T. (2011). Bernoulli Graph Bounds for Generalrandom Graphs. *Sociological Methodology*, 41(1):299–345.

Butts, C. T. (2019). A dynamic process interpretation of the sparse ERGM reference model. *The Journal of Mathematical Sociology*, 43(1):40–57. Publisher: Routledge _eprint: https://doi.org/10.1080/0022250X.2018.1490737.

Butts, C. T. (2020). A dynamic process reference model for sparse networks with reciprocity. *The Journal of Mathematical Sociology*, pages 1–27.

Butts, C. T. and Acton, R. M. (2011). Spatial Modeling of Social Networks. In *The SAGE Handbook of GIS and Society*, pages 222–250. SAGE Publications, Inc., 1 Oliver's Yard, 55 City Road London EC1Y 1SP.

Butts, C. T., Acton, R. M., Hipp, J. R., and Nagle, N. N. (2012). Geographical variability and network structure. *Social Networks*, 34(1):82–100.

Butts, C. T., Petrescu-Prahova, M., and Cross, R. B. (2007). Responder Communication Networks in the World Trade Center Disaster: Implications for Modeling of Communication Within Emergency Settings. *The Journal of Mathematical Sociology*, 31(2):121–147.

Card, D. (2001). Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration. *Journal of Labor Economics*, 19(1):22–64.

Carling, J. (2002). Migration in the age of involuntary immobility: Theoretical reflections and Cape Verdean experiences. *Journal of Ethnic and Migration Studies*, 28(1):5–42.

Carling, J. and Schewel, K. (2018). Revisiting aspiration and ability in international migration. *Journal of Ethnic and Migration Studies*, 44(6):945–963.

Castles, S., Haas, H. d., and Miller, M. J. (2013). *The Age of Migration: International Population Movements in the Modern World*. Palgrave Macmillan.

Cetina, K. K. and Cicourel, A. V., editors (2014). *Advances in Social Theory and Methodology (RLE Social Theory): Toward an Integration of Micro- and Macro-Sociologies*. Routledge, London.

Charyyev, B. and Gunes, M. H. (2019). Complex network of United States migration. *Computational Social Networks*, 6(1):1–28.

Chen, M. K. and Rohla, R. (2018). The effect of partisanship and political advertising on close family ties. *Science*, 360(6392):1020–1024.

Cheng, S. and Park, B. (2020). Flows and Boundaries: A Network Approach to Studying Occupational Mobility in the Labor Market. *American Journal of Sociology*, 126(3):577–631.

Chu, J. Y. (2010). *Cosmologies of Credit: Transnational Mobility and the Politics of Destination in China*. Duke University Press. Google-Books-ID: eKPyp60pOloC.

Clark, W. A. V. (2008). Reexamining the moving to opportunity study and its contribution to changing the distribution of poverty and ethnic concentration. *Demography*, 45(3):515–535.

Cohen, J. E., Roig, M., Reuman, D. C., and GoGwilt, C. (2008). International migration beyond gravity: A statistical model for use in population projections. *Proceedings of the National Academy of Sciences*, 105(40):15269–15274.

Cohen, J. H. and Sirkeci, I. (2011). *Cultures of Migration: The Global Nature of Contemporary Mobility*. University of Texas Press.

Coleman, J. S. (1986). Social Theory, Social Research, and a Theory of Action. *American Journal of Sociology*, 91(6):1309–1335.

Conway, D. (1980). Step-Wise Migration: Toward a Clarification of the Mechanism. *International Migration Review*, 14(1):3–14.

Cooke, T. J. (2013). Internal Migration in Decline. *The Professional Geographer*, 65(4):664–675.

Coulson, N. E. and Grieco, P. L. E. (2013). Mobility and mortgages: Evidence from the PSID. *Regional Science and Urban Economics*, 43(1):1–7.

Cramer, K. J. (2016). *The Politics of Resentment: Rural Consciousness in Wisconsin and the Rise of Scott Walker*. University of Chicago Press, Chicago.

Cranmer, S. J. and Desmarais, B. A. (2011). Inferential Network Analysis with Exponential Random Graph Models. *Political Analysis*, 19(1):66–86.

Crowder, K., Pais, J., and South, S. J. (2012). Neighborhood Diversity, Metropolitan Constraints, and Household Migration. *American Sociological Review*, 77(3):325–353.

Crowder, K., South, S. J., and Chavez, E. (2006). Wealth, Race, and Inter-Neighborhood Migration. *American Sociological Review*, 71(1):72–94.

Czaika, M. and de Haas, H. (2017). The Effect of Visas on Migration Processes. *International Migration Review*, 51(4):893–926.

Dao, M., Furceri, D., and Loungani, P. (2017). Regional Labor Market Adjustment in the United States: Trend and Cycle. *The Review of Economics and Statistics*, 99(2):243–257.

Davis, J. A. and Leinhardt, S. (1972). The Structure of Positive Interpersonal Relations in Small Groups. In *Sociological Theories in Progress*, pages 218–251. Houghton Mifflin, Boston.

de Haas, H. (2010). The Internal Dynamics of Migration Processes: A Theoretical Inquiry. *Journal of Ethnic and Migration Studies*, 36(10):1587–1617.

Dekker, D., Krackhardt, D., and Snijders, T. A. B. (2007). Sensitivity of MRQAP Tests to Collinearity and Autocorrelation Conditions. *Psychometrika*, 72(4):563–581.

DellaPosta, D. (2020). Pluralistic Collapse: The “Oil Spill” Model of Mass Opinion Polarization. *American Sociological Review*, 85(3):507–536.

DellaPosta, D. and Macy, M. (2015). The center cannot hold. Networks, echo chambers, and polarization. In *Order on the Edge of Chaos, Social Psychology and the Problem of Social Order*, pages 86–104. Cambridge University Press, New York.

DellaPosta, D., Shi, Y., and Macy, M. (2015). Why Do Liberals Drink Lattes? *American Journal of Sociology*, 120(5):1473–1511.

DeLuca, S., Wood, H., and Rosenblatt, P. (2019). Why Poor Families Move (And Where They Go): Reactive Mobility and Residential Decisions. *City & Community*, 18(2):556–593.

Desmarais, B. and Cranmer, S. (2012a). Statistical mechanics of networks: Estimation and uncertainty. *Physica A: Statistical Mechanics and its Applications*, 391(4):1865–1876.

Desmarais, B. A. and Cranmer, S. J. (2012b). Statistical Inference for Valued-Edge Networks: The Generalized Exponential Random Graph Model. *PLoS ONE*, 7(1):e30136.

DeWaard, J., Curtis, K. J., and Fussell, E. (2016). Population recovery in New Orleans after Hurricane Katrina: exploring the potential role of stage migration in migration systems. *Population and Environment*, 37(4):449–463.

DeWaard, J., Fussell, E., Curtis, K. J., and Ha, J. T. (2020). Changing spatial interconnectivity during the “Great American Migration Slowdown”: A decomposition of intercounty migration rates, 1990–2010. *Population, Space and Place*, 26(1):e2274.

DeWaard, J. and Ha, J. T. (2019). Resituating relaunched migration systems as emergent entities manifested in geographic structures. *Migration Studies*, 7(1):39–58.

DeWaard, J., Hauer, M., Fussell, E., Curtis, K. J., Whitaker, S. D., McConnell, K., Price, K., Egan-Robertson, D., Soto, M., and Castro, C. A. (2022). User Beware: Concerning Findings from the Post 2011–2012 U.S. Internal Revenue Service Migration Data. *Population Research and Policy Review*, 41(2):437–448.

DeWaard, J., Kim, K., and Raymer, J. (2012). Migration Systems in Europe: Evidence From Harmonized Flow Data. *Demography*, 49(4):1307–1333.

DiPrete, T. A., Gelman, A., McCormick, T., Teitler, J., and Zheng, T. (2011). Segregation in Social Networks Based on Acquaintanceship and Trust. *American Journal of Sociology*, 116(4):1234–83.

Dorsey, C. (2021). America’s Mass Migration Intensifies As ‘Leftugees’ Flee Blue States And Counties For Red. *Forbes*.

Drabek, T. E., Tamminga, H. L., Kilijanek, T. S., and Adams, C. R. (1981). *Managing multiorganizational emergency responses: emergent search and rescue networks in natural disaster and remote area settings*. Univ. Inst. of Behavioral Science.

Durand, J. and Massey, D. S. (2010). New world orders: Continuities and changes in Latin American migration. *The Annals of the American Academy of Political and Social Science*, 630(1):20–52. Publisher: Sage Publications Sage CA: Los Angeles, CA.

Duxbury, S. W. (2023). Micro Effects on Macro Structure in Social Networks. *Sociological Methodology*, page 00811750231209040. Publisher: SAGE Publications Inc.

Dye, T. R. (1990). *American Federalism: Competition Among Governments.* Lexington Books.

Eddelbuettel, D. and Balamuta, J. J. (2018). Extending R with C++: A brief introduction to Rcpp. *The American Statistician*, 72(1):28–36. Publisher: Taylor & Francis.

Eeckhout, J. (2004). Gibrat's Law for (All) Cities. *American Economic Review*, 94(5):1429–1451.

Einstein, A., Podolsky, B., and Rosen, N. (1935). Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? *Physical Review*, 47(10):777–780. Cambridge, UK: Cambridge University Press.

Faust, K. (2011). Animal social networks. In *The SAGE handbook of social network analysis. London: SAGE Publications*, pages 148–66. SAGE Publications Ltd, London.

Fawcett, J. T. (1989). Networks, Linkages, and Migration Systems. *The International Migration Review*, 23(3):671–680.

Fernández-Kelly, P. (2008). The Back Pocket Map: Social Class and Cultural Capital as Transferable Assets in the Advancement of Second-Generation Immigrants. *The ANNALS of the American Academy of Political and Social Science*, 620(1):116–137.

Fischer, C. S. (2002). Ever-More Rooted Americans. *City & Community*, 1(2):177–198.

Fitchen, J. M. (1994). Residential Mobility Among the Rural Poor. *Rural Sociology*, 59(3):416–436. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1549-0831.1994.tb00540.x.

Fossett, M. (2006). Ethnic Preferences, Social Distance Dynamics, and Residential Segregation: Theoretical Explorations Using Simulation Analysis. *The Journal of Mathematical Sociology*, 30(3-4):185–273.

Fowler, J. H. (2006). Connecting the Congress: A Study of Cosponsorship Networks. *Political Analysis*, 14(4):456–487. Publisher: Cambridge University Press.

Freier, L. F. and Holloway, K. (2019). The Impact of Tourist Visas on Intercontinental South-South Migration: Ecuador's Policy of "Open Doors" as a Quasi-Experiment. *International Migration Review*, 53(4):1171–1208.

Frey, W. H. (1979). Central City White Flight: Racial and Nonracial Causes. *American Sociological Review*, 44(3):425–448.

Frey, W. H. (1995a). Immigration and Internal Migration 'Flight' from US Metropolitan Areas: Toward a New Demographic Balkanisation. *Urban Studies*, 32(4-5):733–757.

Frey, W. H. (1995b). Immigration and internal migration "flight": A California case study. *Population and Environment*, 16(4):353–375.

Frey, W. H. (2009). The great American migration slowdown: Regional and metropolitan dimensions. Technical report, Brookings Institution, Washington, DC.

Garip, F. (2008). Social capital and migration: How do similar resources lead to divergent outcomes? *Demography*, 45(3):591–617.

Garip, F. (2019). *On the Move: Changing Mechanisms of Mexico-U.S. Migration*. Princeton University Press.

Garip, F. and Asad, A. L. (2016). Network Effects in Mexico–U.S. Migration: Disentangling the Underlying Social Mechanisms. *American Behavioral Scientist*, 60(10):1168–1193.

Geyer, C. J. and Thompson, E. A. (1992). Constrained Monte Carlo Maximum Likelihood for Dependent Data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 54(3):657–683.

Gimpel, J. G. and Hui, I. S. (2015). Seeking politically compatible neighbors? The role of neighborhood partisan composition in residential sorting. *Political Geography*, 48:130–142.

Gondal, N. (2018). Duality of departmental specializations and PhD exchange: A Weberian analysis of status in interaction using multilevel exponential random graph models (mERGM). *Social Networks*, 55:202–212.

Goodreau, S. M., Kitts, J. A., and Morris, M. (2009). Birds of a feather, or friend of a friend? using exponential random graph models to investigate adolescent social networks. *Demography*, 46(1):103–125.

Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology*, 78(6):1360–1380.

Greenwood, M. J. (1997). Chapter 12 Internal migration in developed countries. In *Handbook of Population and Family Economics*, volume 1, pages 647–720. Elsevier.

Grusky, D. B., Western, B., and Wimer, C. (2011). *The Great Recession*. Russell Sage Foundation.

Hall, B., Limaye, A., and Kulkarni, A. B. (2009). Overview: Generation of Gene Knockout Mice. *Current Protocols in Cell Biology*, 44(1):19.12.1–19.12.17.

Han, X., Xu, Y., Fan, L., Huang, Y., Xu, M., and Gao, S. (2021). Quantifying COVID-19 importation risk in a dynamic network of domestic cities and international countries. *Proceedings of the National Academy of Sciences*, 118(31):e2100201118.

Handcock, M. S. (2003). *Statistical models for social networks: Inference and degeneracy*. na.

Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., and Morris, M. (2008). statnet: Software Tools for the Representation, Visualization, Analysis and Simulation of Network Data. *Journal of Statistical Software*, 24(1):1548–7660.

Hanneke, S., Fu, W., and Xing, E. P. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4:585–605.

Hauer, M. and Byars, J. (2019). IRS county-to-county migration data, 1990–2010. *Demographic Research*, 40:1153–1166.

Hauer, M. E. (2017). Migration induced by sea-level rise could reshape the US population landscape. *Nature Climate Change*, 7(5):321–325.

Henrie, C. J. and Plane, D. A. (2008). Exodus from the California Core: Using Demographic Effectiveness and Migration Impact Measures to Examine Population Redistribution Within the Western United States. *Population Research and Policy Review*, 27(1):43–64.

Herting, J. R., Grusky, D. B., and Rompaey, S. E. V. (1997). The Social Geography of Interstate Mobility and Persistence. *American Sociological Review*, 62(2):267–287.

Hiltzik, M. (2020). California isn't "hemorrhaging" people, but there are reasons for concern. *Los Angeles Times*.

Hinton, G. E. (2002). Training Products of Experts by Minimizing Contrastive Divergence. *Neural Computation*, 14(8):1771–1800.

Hipp, J. R., Butts, C. T., Acton, R., Nagle, N. N., and Boessen, A. (2013). Extrapolative simulation of neighborhood networks based on population spatial distribution: Do they predict crime? *Social Networks*, 35(4):614–625.

Hochschild, A. R. (2018). *Strangers in Their Own Land: Anger and Mourning on the American Right*. The New Press.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association*, 97(460):1090–1098.

Holland, P. W. and Leinhardt, S. (1981). An Exponential Family of Probability Distributions for Directed Graphs. *Journal of the American Statistical Association*, 76(373):33–50.

Huang, P. and Butts, C. T. (2023). Rooted America: Immobility and Segregation of the Intercounty Migration Network. *American Sociological Review*, 88(6):1031–1065.

Huang, P. and Butts, C. T. (2024a). California Exodus? A network model of population redistribution in the United States. *The Journal of Mathematical Sociology*, 48(3):311–339.

Huang, P. and Butts, C. T. (2024b). Parameter estimation procedures for exponential-family random graph models on count-valued networks: A comparative simulation study. *Social Networks*, 76:51–67.

Hummel, R. M., Hunter, D. R., and Handcock, M. S. (2012). Improving Simulation-Based Algorithms for Fitting ERGMs. *Journal of Computational and Graphical Statistics*, 21(4):920–939.

Hunter, D. R., Goodreau, S. M., and Handcock, M. S. (2008a). Goodness of Fit of Social Network Models. *Journal of the American Statistical Association*, 103(481):248–258.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., and Morris, M. (2008b). ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks. *Journal of statistical software*, 24(3):nihpa54860.

Hunter, D. R., Krivitsky, P. N., and Schweinberger, M. (2012). Computational Statistical Methods for Social Network Models. *Journal of Computational and Graphical Statistics*, 21(4):856–882.

Hyatt, H., McEntarfer, E., Ueda, K., and Zhang, A. (2018). Interstate Migration and Employer-to-Employer Transitions in the United States: New Evidence From Administrative Records Data. *Demography*, 55(6):2161–2180.

Hyatt, H. R. and Spletzer, J. R. (2013). The recent decline in employment dynamics. *IZA Journal of Labor Economics*, 2(1):1–21.

Hyvärinen, A. (2006). Consistency of pseudolikelihood estimation of fully visible Boltzmann machines. *Neural Computation*, 18(10):2283–2292. Publisher: MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info ....

Intrator, J., Tannen, J., and Massey, D. S. (2016). Segregation by race and income in the United States 1970–2010. *Social Science Research*, 60:45–60.

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., and Westwood, S. J. (2019). The Origins and Consequences of Affective Polarization in the United States. *Annual Review of Political Science*, 22(1):129–146.

Iyengar, S., Sood, G., and Lelkes, Y. (2012). Affect, Not IdeologyA Social Identity Perspective on Polarization. *Public Opinion Quarterly*, 76(3):405–431.

Jasso, G. (2011). Migration and stratification. *Social Science Research*, 40(5):1292–1336.

Jennissen, R. (2007). Causality Chains in the International Migration Systems Approach. *Population Research and Policy Review*, 26(4):411–436.

Jia, N., Molloy, R., Smith, C., and Wozniak, A. (2023). The Economics of Internal Migration: Advances and Policy Questions. *Journal of Economic Literature*, 61(1):144–180.

Johnson, J. E. and Kleiner, M. M. (2020). Is Occupational Licensing a Barrier to Interstate Migration? *American Economic Journal: Economic Policy*, 12(3):347–373.

Kaplan, G. and Schulhofer-Wohl, S. (2017). Understanding the Long-Run Decline in Interstate Migration. *International Economic Review*, 58(1):57–94.

Kim, K. and Cohen, J. E. (2010). Determinants of International Migration Flows to and from Industrialized Countries: A Panel Data Approach beyond Gravity. *International Migration Review*, 44(4):899–932.

Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social Networks*, 10(4):359–381.

Krackhardt, D. and Porter, L. W. (1986). The snowball effect: Turnover embedded in communication networks. *Journal of Applied Psychology*, 71(1):50–55.

Krackhardt, D. and Stern, R. N. (1988). Informal Networks and Organizational Crises: An Experimental Simulation. *Social Psychology Quarterly*, 51(2):123–140.

Kreager, D. A., Young, J. T., Haynie, D. L., Bouchard, M., Schaefer, D. R., and Zajac, G. (2017). Where "Old Heads" Prevail: Inmate Hierarchy in a Men's Prison Unit. *American Sociological Review*, 82(4):685–718.

Kritz, M. M. and Gurak, D. T. (2001). The impact of immigration on the internal migration of natives and immigrants. *Demography*, 38(1):133–145.

Kritz, M. M., Lim, L. L., Zlotnik, H., and Lim, L. L. L. L. (1992). *International migration systems: a global approach*. Oxford University Press, USA.

Krivitsky, P. N. (2012). Exponential-family random graph models for valued networks. *Electronic journal of statistics*, 6:1100–1128.

Krivitsky, P. N. (2017). Using contrastive divergence to seed Monte Carlo MLE for exponential-family random graph models. *Computational Statistics & Data Analysis*, 107:149–161.

Krivitsky, P. N. and Butts, C. T. (2013). Modeling valued networks with statnet. *The Statnet Development Team*.

Krivitsky, P. N. and Butts, C. T. (2017). Exponential-family Random Graph Models for Rank-order Relational Data. *Sociological Methodology*, 47(1):68–112.

Krivitsky, P. N., Handcock, M. S., and Hunter, D. R. (2012). Package 'ergm. count'. *Journal of Statistics*, 6:1100–1128.

Krivitsky, P. N., Hunter, D. R., Morris, M., and Klumb, C. (2022). ergm 4: Computational Improvements. Number: arXiv:2203.08198 arXiv:2203.08198 [stat].

Kruse, K. M. (2013). White Flight: Atlanta and the Making of Modern Conservatism. In *White Flight*. Princeton University Press.

Krysan, M. and Crowder, K. (2017). *Cycle of Segregation: Social Processes and Residential Stratification*. Russell Sage Foundation.

Lakon, C. M., Hipp, J. R., Wang, C., Butts, C. T., and Jose, R. (2015). Simulating Dynamic Network Models and Adolescent Smoking: The Impact of Varying Peer Influence and Peer Selection. *American Journal of Public Health*, 105(12):2438–2448.

Leal, D. F. (2021). Network Inequalities and International Migration in the Americas. *American Journal of Sociology*, 126(5):1067–1126.

Lee, E. S. (1966). A theory of migration. *Demography*, 3(1):47–57.

Lee, J. and Zhou, M. (2017). Why class matters less for Asian-American academic achievement. *Journal of Ethnic and Migration Studies*, 43(14):2316–2330.

Leszczensky, L. and Pink, S. (2019). What Drives Ethnic Homophily? A Relational Approach on How Ethnic Identification Moderates Preferences for Same-Ethnic Friends. *American Sociological Review*, 84(3):394–419.

Levendusky, M. (2009). *The Partisan Sort: How Liberals Became Democrats and Conservatives Became Republicans*. University of Chicago Press.

Lewis, K. (2013). The limits of racial prejudice. *Proceedings of the National Academy of Sciences*, 110(47):18814–18819.

Lewis, K. (2016). Preferences in the Early Stages of Mate Choice. *Social Forces*, 95(1):283–320.

Lewis, K. and Papachristos, A. V. (2019). Rules of the Game: Exponential Random Graph Models of a Gang Homicide Network. *Social Forces*, 98(4):1829–1858.

Liang, Z., Chunyu, M., Zhuang, G., and Ye, W. (2008). Cumulative Causation, Market Transition, and Emigration from China. *American Journal of Sociology*, 114(3):706–737.

Liang, Z. and Chunyu, M. D. (2013). Migration within China and from China to the USA: The effects of migration networks, selectivity, and the rural political economy in Fujian Province. *Population Studies*, 67(2):209–223.

Lichter, D. T. and Brown, D. L. (2011). Rural America in an Urban Society: Changing Spatial and Social Boundaries. *Annual Review of Sociology*, 37(1):565–592.

Lichter, D. T., Parisi, D., and Taquino, M. C. (2022). Inter-County Migration and the Spatial Concentration of Poverty: Comparing Metro and Nonmetro Patterns. *Rural Sociology*, 87(1):119–143.

Liu, J. M. and Peng, R. J. (2023). Mobility Repertoires: How Chinese Overseas Students Overcame Pandemic-Induced Immobility. *International Migration Review*, page 01979183231170835.

Liu, X., Andris, C., and Desmarais, B. A. (2019). Migration and political polarization in the U.S.: An analysis of the county-level migration network. *PLOS ONE*, 14(11):e0225405.

Lobao, L. and Kelly, P. (2019). Local Governments across the Rural–Urban Continuum: Findings from a Recent National County Government Study. *State and Local Government Review*, 51(4):223–232.

Long, L. (1991). Residential Mobility Differences among Developed Countries. *International Regional Science Review*, 14(2):133–147.

Lu, Y., Liang, Z., and Chunyu, M. D. (2013). Emigration from China in Comparative Perspective. *Social Forces*, 92(2):631–658.

Lubbers, M. J. and Snijders, T. A. B. (2007). A comparison of various approaches to the exponential random graph model: A reanalysis of 102 student networks in school classes. *Social Networks*, 29(4):489–507.

Lubbers, M. J., Verdery, A. M., and Molina, J. L. (2020). Social Networks and Transnational Social Fields: A Review of Quantitative and Mixed-Methods Approaches. *International Migration Review*, 54(1):177–204.

Mabogunje, A. L. (1970). Systems Approach to a Theory of Rural-Urban Migration. *Geographical Analysis*, 2(1):1–18.

Massey, D. and Denton, N. (1988). The dimensions of residential segregation. *Social Forces*, 67(2):281–315.

Massey, D. and Denton, N. A. (1993). *American Apartheid: Segregation and the Making of the Underclass*. Harvard University Press.

Massey, D. S. (1990). Social Structure, Household Strategies, and the Cumulative Causation of Migration. *Population Index*, 56(1):3.

Massey, D. S., Alarcón, R., Durand, J., and González, H. (1990). *Return to Aztlan: The social process of international migration from Western Mexico*. Univ of California Press.

Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., and Pellegrino, A. (1999). *Worlds in Motion: Understanding International Migration at the End of the Millennium*. Clarendon Press.

Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., and Taylor, J. E. (1993). Theories of International Migration: A Review and Appraisal. *Population and Development Review*, 19(3):431–466.

Massey, D. S., Durand, J., and Pren, K. A. (2016). Why Border Enforcement Backfired. *American Journal of Sociology*, 121(5):1557–1600.

Massey, D. S. and Espinosa, K. E. (1997). What's Driving Mexico-U.S. Migration? A Theoretical, Empirical, and Policy Analysis. *American Journal of Sociology*, 102(4):939–999.

Massey, D. S., Goldring, L., and Durand, J. (1994a). Continuities in Transnational Migration: An Analysis of Nineteen Mexican Communities. *American Journal of Sociology*, 99(6):1492–1533.

Massey, D. S., Gross, A. B., and Shibuya, K. (1994b). Migration, Segregation, and the Geographic Concentration of Poverty. *American Sociological Review*, 59(3):425.

Massey, D. S. and Tannen, J. (2018). Suburbanization and segregation in the United States: 1970–2010. *Ethnic and Racial Studies*, 41(9):1594–1611.

Mata-Codesal, D. (2015). Ways of Staying Put in Ecuador: Social and Embodied Experiences of Mobility–Immobility Interactions. *Journal of Ethnic and Migration Studies*, 41(14):2274–2290.

McFarland, D. A., Moody, J., Diehl, D., Smith, J. A., and Thomas, R. J. (2014). Network Ecology and Adolescent Social Structure. *American Sociological Review*, 79(6):1088–1121.

McMahan, P. and McFarland, D. A. (2021). Creative Destruction: The Structural Consequences of Scientific Curation. *American Sociological Review*, 86(2):341–376.

McMillan, C. (2019). Tied Together: Adolescent Friendship Networks, Immigrant Status, and Health Outcomes. *Demography*, 56(3):1075–1103.

McMillan, C. (2022). Worth the weight: Conceptualizing and measuring strong versus weak tie homophily. *Social Networks*, 68:139–147.

McMillan, C. (2024). Connecting the dots: Using social network analysis to disentangle the factors driving international migration. *International Migration*, n/a.

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27(1):415–444.

Mele, A. (2017). A Structural Model of Dense Network Formation. *Econometrica*, 85(3):825–850. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA10400.

Miller, J. H. and Page, S. (2009). *Complex Adaptive Systems: An Introduction to Computational Models of Social Life.* Princeton University Press. Publication Title: Complex Adaptive Systems.

MIT Election Data and Science Lab (2018). County Presidential Election Returns 2000-2016. Retrieved September 29, 2020 (`https://doi.org/10.7910/DVN/VOQCHQ`).

Molloy, R., Smith, C. L., and Wozniak, A. (2011). Internal Migration in the United States. *Journal of Economic Perspectives*, 25(3):173–196.

Molloy, R., Smith, C. L., and Wozniak, A. (2017). Job Changing and the Decline in Long-Distance Migration in the United States. *Demography*, 54(2):631–653.

Monras, J. (2018). Economic Shocks and Internal Migration. *CEPR Discussion Paper No. DP12977.*

Moody, J. (2001). Race, School Integration, and Friendship Segregation in America. *American Journal of Sociology*, 107(3):679–716.

Morris, M., Handcock, M. S., and Hunter, D. R. (2008). Specification of Exponential-Family Random Graph Models: Terms and Computational Aspects. *Journal of statistical software*, 24(4):1548–7660.

Mouw, T., Chavez, S., Edelblute, H., and Verdery, A. (2014). Binational Social Networks and Assimilation: A Test of the Importance of Transnationalism. *Social Problems*, 61(3):329–359.

Mueller, J. T. and Gasteyer, S. (2023). The ethnically and racially uneven role of water infrastructure spending in rural economic development. *Nature Water*, 1(1):74–82. Number: 1 Publisher: Nature Publishing Group.

Mummolo, J. and Nall, C. (2016). Why Partisans Do Not Sort: The Constraints on Political Segregation. *The Journal of Politics*, 79(1):45–59. Publisher: The University of Chicago Press.

National Bureau of Economic Research (2016). County Distance Database. Retrieved September 29, 2020 (`https://data.nber.org/data/county-distance-database.html`).

Nogle, J. M. (1994). The Systems Approach to International Migration: An Application of Network Analysis Methods. *International Migration*, 32(2):329–342.

Nowicki, K. and Snijders, T. A. B. (2001). Estimation and Prediction for Stochastic Blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087.

Opsahl, T., Agneessens, F., and Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3):245–251.

Palloni, A., Massey, D., Ceballos, M., Espinosa, K., and Spittel, M. (2001). Social Capital and International Migration: A Test Using Information on Family Networks. *American Journal of Sociology*, 106(5):1262–1298.

Papachristos, A. V., Hureau, D. M., and Braga, A. A. (2013). The Corner and the Crew: The Influence of Geography and Social Networks on Gang Violence. *American Sociological Review*, 78(3):417–447.

Partridge, M. D., Rickman, D. S., Olfert, M. R., and Ali, K. (2012). Dwindling U.S. internal migration: Evidence of spatial equilibrium or structural shifts in local labor markets? *Regional Science and Urban Economics*, 42(1-2):375–388.

Patuelli, R. (2016). Spatial Autocorrelation and Spatial Interaction. In *Encyclopedia of GIS*, pages 1–7. Springer.

Paul, A. M. (2011). Stepwise International Migration: A Multistage Migration Pattern for the Aspiring Migrant. *American Journal of Sociology*, 116(6):1842–86.

Paul, A. M. (2017). *Multinational Maids: Stepwise Migration in a Global Labor Market.* Cambridge University Press.

Piore, M. J. (2018). The Dual Labor Market: Theory and Implications. In *Social Stratification*, pages 629–640. Routledge, 4 edition.

Plane, D. A. (1999). Migration Drift. *The Professional Geographer*, 51(1):1–11.

Plane, D. A., Henrie, C. J., and Perry, M. J. (2005). Migration up and down the urban hierarchy and across the life course. *Proceedings of the National Academy of Sciences*, 102(43):15313–15318.

Plantinga, A. J., Détang-Dessendre, C., Hunt, G. L., and Piguet, V. (2013). Housing prices and inter-urban migration. *Regional Science and Urban Economics*, 43(2):296–306.

Poot, J., Alimi, O., Cameron, M. P., and Maré, D. C. (2016). The Gravity Model of Migration: The Successful Comeback of an Ageing Superstar in Regional Science. *SSRN Electronic Journal*.

Preuhs, R. R. (1999). State Policy Components of Interstate Migration in the United States. *Political Research Quarterly*, 52(3):527–549.

Preuhs, R. R. (2020). Pack Your Politics! Assessing the Vote Choice of Latino Interstate Migrants. *The Journal of Race, Ethnicity, and Politics*, 5(1):130–165.

Putnam, R. D. (2000). *Bowling Alone: The Collapse and Revival of American Community.* Simon and Schuster.

Quillian, L. (2015). A Comparison of Traditional and Discrete-Choice Approaches to the Analysis of Residential Mobility and Locational Attainment. *The ANNALS of the American Academy of Political and Social Science*, 660(1):240–260.

Ravenstein, E. G. (1885). The Laws of Migration. *Journal of the Royal Statistical Society*, 48(2):167–235.

Raymer, J. and Rogers, A. (2007). Applying Model Migration Schedules to Represent Age-Specific Migration Flows. In Raymer, J. and Willekens, F., editors, *International Migration in Europe*, pages 175–192. John Wiley & Sons, Ltd, Chichester, UK.

Rees, P., Bell, M., Kupiszewski, M., Kupiszewska, D., Ueffing, P., Bernard, A., Charles-Edwards, E., and Stillwell, J. (2017). The Impact of Internal Migration on Population Redistribution: an International Comparison. *Population, Space and Place*, 23(6):e2036.

Riddell, J. B. and Harvey, M. E. (1972). The Urban System in the Migration Process: An Evaluation of Step-Wise Migration in Sierra Leone. *Economic Geography*, 48(3):270.

Robins, G., Pattison, P., and Wasserman, S. (1999). Logit models and logistic regressions for social networks: III. Valued relations. *Psychometrika*, 64(3):371–394.

Rogers, A. and Castro, L. J. (1981). *Model Migration Schedules*. International Institute for Applied Systems Analysis, Laxenburg.

Ryo, E. (2013). Deciding to Cross: Norms and Economics of Unauthorized Migration. *American Sociological Review*, 78(4):574–603.

Sakoda, J. M. (1971). The checkerboard model of social interaction. *The Journal of Mathematical Sociology*, 1(1):119–132.

Schelling, T. C. (1969). Models of Segregation. *American Economic Review*, 59(2):483–493.

Schelling, T. C. (1978). *Micromotives and Macrobehavior*. W. W. Norton & Company.

Schewel, K. (2020). Understanding Immobility: Moving Beyond the Mobility Bias in Migration Studies. *International Migration Review*, 54(2):328–355.

Schildkraut, D. J., Jiménez, T. R., Dovidio, J. F., and Huo, Y. J. (2019). A Tale of Two States: How State Immigration Climate Affects Belonging to State and Country among Latinos. *Social Problems*, 66(3):332–355.

Schiller, N. G., Basch, L., and Blanc, C. S. (1995). From Immigrant to Transmigrant: Theorizing Transnational Migration. *Anthropological Quarterly*, 68(1):48–63.

Schmid, C. S. and Desmarais, B. A. (2017). Exponential random graph models with big networks: Maximum pseudolikelihood estimation and the parametric bootstrap. In *2017 IEEE International Conference on Big Data (Big Data)*, pages 116–121.

Schroeder, J. P. and Pacas, J. D. (2021). Across the Rural–Urban Universe: Two Continuous Indices of Urbanization for U.S. Census Microdata. *Spatial Demography*, 9(1):131–154.

Sharkey, P. (2015). Geographic Migration of Black and White Families Over Four Generations. *Demography*, 52(1):209–231.

Shryock, H. S., Siegel, J. S., and Larmon, E. A. (1973). *The Methods and Materials of Demography*. U.S. Bureau of the Census.

Simpson, S. L., Bowman, F. D., and Laurienti, P. J. (2013). Analyzing complex functional brain networks: Fusing statistics and network science to understand the brain. *Statistics surveys*, 7:1–36.

Smith, C. M. and Papachristos, A. V. (2016). Trust Thy Crooked Neighbor: Multiplexity in Chicago Organized Crime Networks. *American Sociological Review*, 81(4):644–667.

Smith, J. A., McPherson, M., and Smith-Lovin, L. (2014). Social Distance in the United States: Sex, Race, Religion, Age, and Education Homophily among Confidants, 1985 to 2004. *American Sociological Review*, 79(3):432–456.

Snijders, T. A. B. (2002). Markov Chain Monte Carlo Estimation of Exponential Random Graph Models. *Journal of Social Structure*, 3(2):1–40.

Song, S. (2021). Study shows California exodus, with more people leaving the state despite the pandemic. *KTVU FOX 2*.

Sparrowe, R. T. and Liden, R. C. (1997). Process and Structure in Leader-Member Exchange. *Academy of Management Review*, 22(2):522–552.

Spring, A., Mulder, C. H., Thomas, M. J., and Cooke, T. J. (2021). Migration after union dissolution in the United States: The role of non-resident family. *Social Science Research*, 96:102539.

Srivastava, S. B. and Banaji, M. R. (2011). Culture, Cognition, and Collaborative Networks in Organizations. *American Sociological Review*, 76(2):207–233.

Steinbeck, J. (1939). *The Grapes of Wrath*. Penguin.

Stockdale, A. and Haartsen, T. (2018). Editorial introduction: Putting rural stayers in the spotlight. *Population, Space and Place*, 24(4):e2124.

Strauss, D. and Ikeda, M. (1990). Pseudolikelihood Estimation for Social Networks. *Journal of the American Statistical Association*, 85(409):204–212.

Tam Cho, W. K., Gimpel, J. G., and Hui, I. S. (2013). Voter Migration and the Geographic Sorting of the American Electorate. *Annals of the Association of American Geographers*, 103(4):856–870.

Tan, L. S. L. and Friel, N. (2020). Bayesian Variational Inference for Exponential Random Graph Models. *Journal of Computational and Graphical Statistics*, 29(4):910–928. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10618600.2020.1740714.

Thomas, L. J., Huang, P., Yin, F., Xu, J., Almquist, Z. W., Hipp, J. R., and Butts, C. T. (2022). Geographical patterns of social cohesion drive disparities in early COVID infection hazard. *Proceedings of the National Academy of Sciences*, 119(12):e2121675119.

Tiebout, C. M. (1956). A Pure Theory of Local Expenditures. *Journal of Political Economy*, 64(5):416–424.

Tocqueville, A. (1834). *Democracy in America: And Two Essays on America*. Penguin UK.

Todaro, M. P. (1976). *Internal migration in developing countries*. International Labour Office, Genève, Switzerland.

Tolnay, S. E. (2003). The African American "Great Migration" and Beyond. *Annual Review of Sociology*, 29(1):209–232.

Treyz, G. I., Rickman, D. S., Hunt, G. L., and Greenwood, M. J. (1993). The Dynamics of U.S. Internal Migration. *The Review of Economics and Statistics*, 75(2):209–214.

Ulibarri, N. and Scott, T. A. (2017). Linking Network Structure to Collaborative Governance. *Journal of Public Administration Research and Theory*, 27(1):163–181.

U.S. Census Bureau (2011a). American Community Survey 5-Year Estimates 2006-2010. Retrieved September 29, 2020 (`https://data.census.gov/`).

U.S. Census Bureau (2011b). United States Census 2010. Retrieved September 29, 2020 (`https://data.census.gov/`).

U.S. Census Bureau (2013). Census Regions and Divisions of the United States. Retrieved September 29, 2020 (`https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf`).

U.S. Census Bureau (2019). County-to-county migration flows: 2011-2015 ACS. Retrieved September 29, 2020 (`https://www.census.gov/data/tables/2015/demo/geographic-mobility/county-to-county-migration-2011-2015.html`).

van Duijn, M. A. J., Gile, K. J., and Handcock, M. S. (2009). A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models. *Social Networks*, 31(1):52–62.

Vega Yon, G. G., Slaughter, A., and de la Haye, K. (2021). Exponential random graph models for little networks. *Social Networks*, 64:225–238.

Verdery, A. M., Mouw, T., Edelblute, H., and Chavez, S. (2018). Communication flows and the durability of a transnational social field. *Social Networks*, 53:57–71.

Vogel, G. (2007). A Knockout Award in Medicine. *Science*, 318(5848):178–179.

von Reichert, C., Cromartie, J. B., and Arthun, R. O. (2014a). Impacts of Return Migration on Rural U.S. Communities. *Rural Sociology*, 79(2):200–226.

von Reichert, C., Cromartie, J. B., and Arthun, R. O. (2014b). Reasons for Returning and Not Returning to Rural U.S. Communities. *The Professional Geographer*, 66(1):58–72.

Vu, D. Q., Hunter, D. R., and Schweinberger, M. (2013). Model-based Clustering of Large Networks. *The annals of applied statistics*, 7(2):1010.

Vögtle, E. and Windzio, M. (2022). The ‘Global South’ in the transnational student mobility network. Effects of institutional instability, reputation of the higher education systems, post-colonial ties, and culture. *Globalisation, Societies and Education*, pages 1–19.

Wainwright, M. J. and Jordan, M. I. (2008). Graphical Models, Exponential Families, and Variational Inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305. Publisher: Now Publishers, Inc.

Waldinger, R. (2013). Immigrant transnationalism. *Current Sociology*, 61(5-6):756–777.

Waldinger, R. and Fitzgerald, D. (2004). Transnationalism in Question. *American Journal of Sociology*, 109(5):1177–1195. Publisher: The University of Chicago Press.

Wallace, T. and Karra, K. (2020). The True Colors of America’s Political Spectrum Are Gray and Green. *The New York Times*.

Wallerstein, I. (2011). *The Modern World-System I: Capitalist Agriculture and the Origins of the European World-Economy in the Sixteenth Century*. University of California Press.

Wang, C., Hipp, J. R., Butts, C. T., Jose, R., and Lakon, C. M. (2016). Coevolution of adolescent friendship networks and smoking and drinking behaviors with consideration of parental influence. *Psychology of Addictive Behaviors*, 30(3):312–324.

Wang, P., Robins, G., and Pattison, P. (2009). PNet: Program for the Simulation and Estimation of Exponential Random Graph (p*) Models.

Ward, M. D., Ahlquist, J. S., and Rozenas, A. (2013). Gravity's Rainbow: A dynamic latent space model for the world trade network. *Network Science*, 1(1):95–118. Publisher: Cambridge University Press.

Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p*. *Psychometrika*, 61(3):401–425.

Weber, M. (1922). *Wirtschaft und Gesellschaft: Grundriß der Verstehenden Soziologie*. Duncker & Humblot.

White, M. J. and Liang, Z. (1998). The effect of immigration on the internal migration of the native-born population, 1981–1990. *Population Research and Policy Review*, 17(2):141–166.

Wilson, D. S., O'Brien, D. T., and Sesma, A. (2009). Human prosociality from an evolutionary perspective: variation and correlations at a city-wide scale. *Evolution and Human Behavior*, 30(3):190–200.

Wimmer, A. and Lewis, K. (2010). Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook. *American Journal of Sociology*, 116(2):583–642.

Windzio, M. (2018). The network of global migration 1990–2013. *Social Networks*, 53:20–29.

Windzio, M., Teney, C., and Lenkewitz, S. (2019). A network analysis of intra-EU migration flows: how regulatory policies, economic inequalities and the network-topology shape the intra-EU migration space. *Journal of Ethnic and Migration Studies*, 47(5):951–969.

Woldoff, R. A. (2011). *White Flight/Black Flight: The Dynamics of Racial Change in an American Neighborhood*. Cornell University Press.

Wright, R. A., Ellis, M., and Reibel, M. (1997). The Linkage between Immigration and Internal Migration in Large Metropolitan Areas in the United States. *Economic Geography*, 73(2):234–254.

Xie, Y. and Zhang, C. (2019). The long-term impact of the Communist Revolution on social stratification in contemporary China. *Proceedings of the National Academy of Sciences*, 116(39):19392–19397.

Yin, F., Phillips, N. E., and Butts, C. T. (2019). Selection of Exponential-Family Random Graph Models via Held-Out Predictive Evaluation (HOPE). preprint, arXiv.

Zhou, M. (1992). *Chinatown: The Socioeconomic Potential of an Urban Enclave.* Temple University Press.

Zipf, G. K. (1946). The P1 P2/D Hypothesis: On the Intercity Movement of Persons. *American Sociological Review*, 11(6):677–686.

Zipf, G. K. (1949). *Human behavior and the principle of least effort.* Addison-Wesley Press, Oxford, England.