SCALABLE AGENT-BASED MODELING OF FORCED MIGRATION

by

Melonie K Richey
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Computational Social Science

Committee:

______________________________	Dr. Hamdi Kavak, Committee Chair
______________________________	Dr. Andrew Crooks, Committee Member
______________________________	Dr. William G. Kennedy, Committee Member
______________________________	Dr. Robert Simon, Committee Member
______________________________	Dr. Robert Axtell, Committee Member
______________________________	Dr. Jason Kinser, Department Chairperson
______________________________	Dr. Donna M. Fox, Associate Dean, Office of Student Affairs & Special Programs, College of Science
______________________________	Dr. Fernando Miralles-Wilhelm, Dean, College of Science
Date: __________________________	Fall Semester 2020 George Mason University Fairfax, VA

Scalable Agent-Based Modeling of Forced Migration

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Melonie K Richey
Master of Science
Mercyhurst University, 2014
Bachelor of Arts
University of Florida, 2011

Director: Hamdi Kavak, Professor
Department of Computational and Data Sciences

Fall Semester 2020
George Mason University
Fairfax, VA

DEDICATION

"Think at last we have not reached conclusion." – T.S. Eliot

ACKNOWLEDGEMENTS

I would like to thank the many friends and colleagues that made this all possible. It was not trivial and would have amounted to far less without the contributions of many. A special thanks to Mr. Zachary "Moose" Mostowsky for performing code reviews and assisting with model verification. His skills in the science of software ninjary are unmatched.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ABBREVIATIONS

Agent-based model .... ABM
Area of Interest .... AOI
Amazon Web Services...AWS
Armed Conflict Location and Event Data Project .... ACLED
Central African Republic .... CAR
Central Processing Unit ....CPU
Computational Social Science .... CSS
Democratic Republic of the Congo.. DRC
Genetic Algorithm ....GA
Global Interpreter Lock...GIL
Google Cloud Platform ....GCP
Graphical Processing Unit ....GPU
High Performance Computing ....HPC
Humanitarian Data Exchange ....HDX
Internally Displaced Person .... IDP
Interservice/Industry Training, Simulation and Education..................................... I/ITSEC
Machine Learning ....ML
Mean Absolute Error..MAE
Mean Absolute Scaled Error .... MASE
Message Passing Interface .... MPI
Mean Root Error ....MRE
Mean Root Square Error ....MRSE
National Geospatial-Intelligence Agency ....NGA
Normalized Mean Absolute Error..NMAE
Neural Network...NN
Overview, Design concepts and Details ....ODD
Social Network Theory .... SNT
Spatial Interaction Model... SIM
United States Agency for International Development .... USAID
United Nations High Commissioner for Refugees ....UNHCR
United Nationals Office for the Coordination of Humanitarian Affairs.............. UNOCHA
Virtual Machine .... VM

ABSTRACT

SCALABLE AGENT-BASED MODELING OF FORCED MIGRATION

Melonie K Richey, Ph.D

George Mason University, 2020

Dissertation Director: Dr. Hamdi Kavak

Migration studies have a long history in sociology and the social science though the discipline has matured in notable ways over the past several decades. One such way is the attention that has been paid to forced migration resulting from security and conflict events worldwide. The study of forced migration is distinct from the study of voluntary migration, the topic of most research in migration studies since the late 19th century. Another way the field has matured is in the application of computational modeling and simulation methods to the problem domain to augment or complement theoretical or statistical analysis of migration. Despite recent advancements in these two areas, there remains a dearth of research around computational modeling and simulation methods as applied to the study of forced migration. There are many gaps the scientific community of sociologists and computational social scientists must fill before empirical models can be generalized and used for predictive purposes by aid organizations to make decisions about the allocation of resources in response to forced migration events worldwide. For instance, many computational models have been applied to the modeling of voluntary migration but far fewer to forced migration. Of the models developed for forced

migration, many are only theoretical, few are empirical, and only one is designed to run at a scale applicable to ongoing forced migration events worldwide with tens of millions of migrants but does not consider social networks.

The research presented in this dissertation fills some of these gaps by addressing limitations in theoretical knowledge and computational methodology with the application of an agent-based model to a real-world forced migration case study – forced Syrian migration into neighboring Turkey. This research makes the case that agent-based modeling is the appropriate social simulation approach to take for forced migration modeling and discusses recent developments in forced migration theory that have yet to be applied in empirical computational modeling contexts, including social networks. This research also addresses the scale issue, demonstrating an agent-based simulation that runs with up to 25M agents and applying this simulation to a real-world event with 4-6M at-risk persons. The following chapters summarize a two-fold contribution to the sociology, social science, and computational modeling scientific communities: (1) the first empirically tested agent-based simulation to model forced migration that considers migrant social networks and (2) methodological advances to the state-of-the-art of computational forced migration modeling and publicly available computational tools and methods to facilitate future research for researchers in this domain.

1. INTRODUCTION

What follows is an introductory chapter to this dissertation on the use of agent-based modeling and simulation techniques to model and predict forced migration. It begins with an overview of the topic and motivation for the research in Section 1.1. Section 1.2 follows by presenting three distinct research questions. Section 1.3 briefly outlines my approach to investigating these three research questions, and Section 1.4 concludes this introductory chapter with a review of the contributions the work presented herein contributes to both social science and the computational social sciences (CSS).

1.1 Overview & Purpose

Migration studies have a long history in the field of sociology dating back to the late 19th century when the German cartographer Ernst Georg Ravenstein first published his 11 laws of migration (Ravenstein, 1885). Since that time, migration studies have garnered substantial attention in the analytic and research communities, even more in the 21st century given the prominence of displacement events worldwide over the last several decades. As of this writing, the amount of forcibly displaced persons worldwide is now a staggering 80 million according to the United Nations High Commissioner for Refugees (UNHCR, 2020a). Computational methods to model and predict these events are likewise increasing in prevalence and complexity (Klabunde & Willekens, 2016). Academic groups, interagency aid organizations, and migration researchers are devising analytic methods and computational models to better understand and predict the movement of refugees (Hebert, Perez, & Harati, 2018; Frydenlund & De Kock, 2020), asylum-seekers (Hattle, 2016; Langley et al., 2016), internally displaced persons (IDPs) (Frydenlund et al., 2018; Naude, 2010b), returnees (Biondo, Pluchino, & Rapisarda, 2012; Haug, 2008; Rehm, 2012; Cassarino, 2004), and stateless persons worldwide (Disney et al., 2015).

In 2020, the Syrian Arab Republic alone hosts 6 million IDPs (UNHCR, 2020c) while almost 4 million of its citizens seek refuge in neighboring Turkey, 800,000 in Lebanon, and 650,000 in Jordan (UNHCR, 2020d). Equally tragic though lesser magnitude circumstances can be found in places such as Burundi, Nigeria, Central African Republic (CAR), South Sudan, Democratic Republic of the Congo (DRC), Venezuela, Iraq, and Yemen (UNHCR, 2020e). I have personally walked through suburbs of Turkish cities Istanbul, Ankara, and Kayseri in the early days of the Syrian conflict, speaking with refugees, asylum-seekers, and stateless persons or persons without status about their migration experiences, their present locations, and their intentions to return to their country of origin. The stories were equally as revealing as they were heartbreaking, and all the more reason to continue this vein of research for provisioning aid to at-risk and stateless populations. I have separately worked with the National Geospatial-Intelligence Agency (NGA) in applying this type of research to populations in CAR (Richey, 2014b) where displacement occurred due to a type of religious genocide. My personal experiences in both places served as motivation to study migration – specifically, forced migration – in an effort to advance the body of knowledge around forced migration modeling and develop computational methods that aid organizations could one day use to provision aid more efficiently and assist these at-risk populations.

Previous research around migration has focused largely on voluntary migration (Stillwell, 1978; Sarra & Signore, 2010; Fotheringham & O'Kelly, 1989; Goodchild & Smith, 1980) and international migration patterns over time (Poot et al., 2016; Moore & Shellman, 2007; Karemera, Oguledo, & Davis., 2000). These efforts have relied heavily on macro-level statistical modeling from largely an economic vantage point (Harris & Todaro, 1970; Lee, 1966; Massey, 1993; Calvo, 1978; Espindola, Silveria, & Penna, 2006; Bergstrand, 1985; Lewer & Van Den Berg, 2008). Out of this vein of study came many spatial interaction models (SIM) such as the gravity model (Wilson, 1970; Pumain et al., 1995; LeSage & Fischer, 2008). While SIMs appropriately model macro-level migration trends, they are incapable of capturing decisions at the individual level or modeling micro-scale phenomena while providing a construct to observe emergent aggregate behavior from the bottom-up (Gulden, Harrison, & Crooks, 2011). Additionally, much of the existing migration theory has been formulated around the drivers of voluntary migration, not forced migration. While many of the factors that influence migration may be similar across both types of migration, the weighting of these factors will differ (Richmond, 1993; Bohra-Mishra & Massey, 2011; Moore & Shellman, 2006; Docquier, Peri, & Ruyssen, 2014; Dorigo & Tobler, 1983).

The research community and aid organizations are left with a dearth of understanding and applicable modeling techniques around the drivers, patterns, and context of forced migration and the global displacement events referenced above. It was over a decade prior that Scott Edwards encouraged the research community to continue developing and applying ever more advanced computational techniques to the forced migration domain (Edwards, 2008). Since then, throughout the past decade, a collection of one-off attempts have been made by researchers to model forced migration using agent-based models (ABMs) (Frydenlund & De Kock, 2020; Hattle, 2016; Hebert, Perez, & Harati, 2018; Frydenlund et al., 2018; Sokolowski, Banks, & Hayes, 2014; Sokowlowski et al., 2014; Frydenlund et al., 2018) and to understand more thoroughly its complexity in theoretical forms of inquiry before models are even developed (Gray, Hilton, & Bijak, 2017; Collins & Frydenlund, 2016). ABMs are a more attractive modeling alternative for forced migration as they provide a bottom-up approach to modeling migrant decision-making at the individual level without prior specification of system-level trends. Despite the diverse applications mentioned above, there remains only one large-scale empirical application of ABMs to forced migration – the FLEE model that can represent millions of migrants (Suleimenova, Bell, & Groen, 2017; Groen, 2018). Applications of the FLEE model, while realistically scalable, have structural and design features that limit or prevent its applicability to some Areas of Interest (AOI), Therefore, it cannot model all forced migration scenarios and also limits its applicability to short term migration scenarios. Additionally, the single empirical ABMs developed for forced migration does not consider social networks. Recent research has shown that social networks greatly influence the decision making of refugees (Dekker et al., 2018; Borkert, Fisher, & Yafi, 2018; Hinsch & Bijak, 2019; Maitland & Xu, 2016).

In short, though substantive contributions have been made to the domain of forced migration in recent years, comprehensive understanding of and computational inquiry into the complexity of migrant decision-making in uncertain conditions, the influence of social factors and technology, and the ultimate drivers of movement as a result of conflict and security events remain to be challenging topics. This is the motivation for the research presented in this dissertation – to develop an empirical computational model that advances our understanding of forced migration to include the consideration of social networks. open-sourceAmong possible computational modeling techniques, which are discussed in Chapter 2, agent-based modeling is a suitable modeling paradigm because it can be used to explicitly represent forced migration populations' decision-making mechanisms at the individual level inclusive of social factors. Finally, there are over 600,000 refugees from CAR and an equal number of IDPs, over 2M refugees from South Sudan, over 1.8M refugees from the Sahel, 4.5M from Venezuela, and, of course, 4M in Turkey (UNHCR, 2020e). The magnitude of these events necessitates scalable computational modeling capabilities that appropriately capture the sizable number of migrants. These gaps in the forced migration research and CSS communities are the gaps that this dissertation intends to fill.

1.2 Research Questions

In satisfaction of the objective mentioned above, there are three primary research questions in this dissertation. They are written from the perspective of the CSS modeler and social scientist who desires to use the outcomes of this research to further their inquiry into forced migration. These questions address fundamental research gaps around forced migration modeling, the application of ABMs to these contexts at scale, and also the most necessary and least understood aspect of migration – social influence.

RQ1 To what extent does the consideration of social networks in forced migration models improve model accuracy?

This research question represents the newest and least explored area of forced migration modeling endeavors. To date, no existing large-scale forced migration model includes social networks explicitly though numerous studies advocate for or lend credence to their inclusion (Reinhardt et al., 2019; Hinsch & Bijak, 2019; Dekker et al., 2018; Simon, 2019; Frydenlund & De Kock, 2020; Borkert, Fisher, & Yafi, 2018; Gray, Hilton, & Bijak, 2017; Klabunde & Willekens, 2016; Maitland & Xu, 2015). A few studies include social networks in models of voluntary migration with encouraging results (Al-Khulaidy & Swartz, 2020; Epstein & Gang, 2006; Tranos et al., 2015; Havinga & Bocker, 1999; Robinson & Segrott, 2002). This research question exists to investigate a significant gap in the CSS literature in the domain of forced migration modeling and leverage new theory surrounding the influence social networks have on forced migrant decision-making.

RQ2 How can ABMs of forced migration be designed and developed at scale to facilitate further investigation?

Building off the previous question, computational techniques will be designed and developed during the course of the research that RQ1 necessitates. The purpose of this research question is to develop and provision computational resources and apply lessons learned in data processing, model design, and parallel computing for the model that will be developed in response to RQ1. These resources will be shared with fellow researchers in support of future forced migration modeling efforts. The FLEE model framework (Suleimenova & Groen, 2017) contributes the first and only truly generalizable and scalable agent-based forced migration modeling framework to the research community, but it has several limitations in its design and application that call into question the utility of the computational methodologies provided within the framework for future research. Most notably, its simulation environment is limiting, and it does not include any way of modeling social influence. While RQ1 addresses the limitation concerned with social influence, RQ2 will address computational tools and techniques that may be used to create more representative simulation environments with less manual effort in support of future modeling efforts.

RQ3 How can the model(s) developed in response to RQ1 and RQ2 be applied to predict where refugee populations are likely to move during a forced migration event in the context of a case study?

This final research question is concerned with the application of the methods designed and developed under the previous two research questions to a case study. Statistical and computational modeling efforts applied to voluntary migration over the years have found predictive success and also yielded great insight into the drivers – economic and otherwise – of this phenomenon. The same cannot be said for forced migration. As mentioned above and discussed more thoroughly in Chapter 2, these traditional techniques are not best suited to modeling reactive, forced, or conflict-induced migration. To address these concerns, this question explores the application of a new

ABM of forced migration to a case study to test and validate it empirically against real-world conditions. To date, the ABMs of forced migration which include social network components are theoretical (Reinhardt et al., 2019; Hisch & Bijak, 2019; Perez, & Harati, 2018; Collins & Frydenlund, 2016) and none have been empirically tested – this study is the first.

1.3 Approach

To address the research objective and three research questions, a four-phased approach is taken: Define, Review, Develop, and Apply. These phases are based loosely on the approach to social simulation defined in Gilbert & Troitzsch, (2005) in that all steps are included yet re-imagined in a workflow that supports structured inquiry into the research questions outlined above. Subordinate phases capture relevant steps but, at a high level, the approach taken is to define the research objective, review relevant literature, design and develop an ABM, and apply that ABM to a relevant case study as depicted in Figure 1.

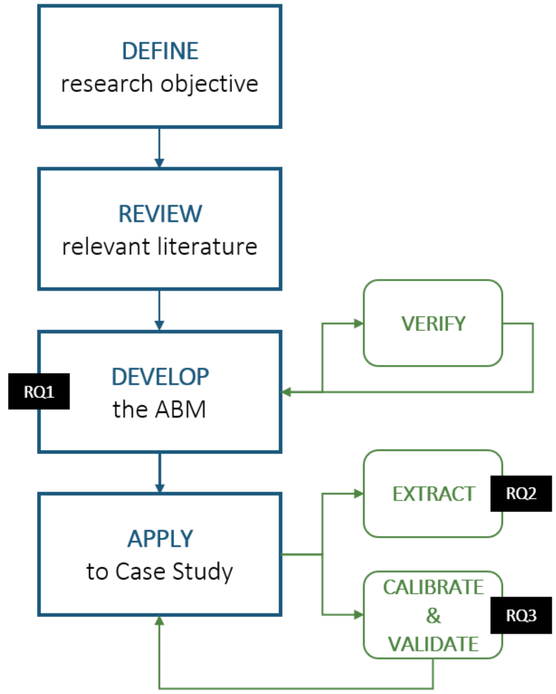


Figure 1 Methodology for approaching the research in this dissertation.

The response to RQ1 exists within the Develop phase. During this phase, the ABM is also verified to ensure internal consistency. The final phase is the Apply phase, where the model is validated through application to a relevant case study in response to RQ3. Generalizable computational techniques and components are also extracted from the Apply phase in support of RQ2 to provide support to subsequent modeling efforts within the scientific community. This methodology is described in greater detail in Chapters 3 and 4.

1.4 Contributions to Social Science

The research presented herein contributes in several distinct ways to both social science and CSS. It advances upon the state of the art in computational modeling of a sociocultural phenomenon that is only growing in importance as conflict and security events worldwide continue to affect the human geographic landscape of our globe. The eventual implications of this type of research for aid organizations such as the United States Agency for International Development (USAID), the United Nations High Commissioner for Refugees (UNHCR), and the United Nations Office for the Coordination of Humanitarian Affairs (UNOCHA) are clear. Accurate decision support tools such as generalizable ABMs will allow these organizations to model where refugees are likely to congregate so that new refugee camps may be considered and established, or alternative resources provisioned in support of local governments. More specifically, concerning the contribution to the scientific community, this research makes contributions in the following distinct areas:

Contribution 1: the first large-scale ABM of forced migration to consider migrant social networks

Current agent-based modeling and simulation efforts for predicting the movement of forced migrants are limited to one-off, non-generalizable models, only one of which is validated empirically and designed to computationally scale in support of mass migration events. The FLEE model is the most robust and empirically validated model to date tested in five separate AOIs: Burundi, Mali, CAR, Iraq, and South Sudan (Suleimenova, Bell, & Groen, 2017). Despite its claimed generalizability, however, it does not include a social component, which I have shown is essential to the modeling of forced migration in the previous section. The single most prominent contribution of this research is a model of forced migration that is performant in its target AOI while introducing the consideration of social networks in the model design.

Social network mechanics can be challenging to include in simulations at scale for two primary reasons: 1) authoritative social network data is lacking for migrant social networks and 2) large graphs and networks can be computationally intensive to include in modeling efforts. This research has overcome the second challenge by including a proxy social network in the presented model and parallelizing the model to the extent that intensive computation is possible in cloud environments and on virtual machines at the scale of tens of millions of refugee agents. The research has addressed the first challenge by using a stylized implementation of the social network. The result of this is a case study which reinforces the preliminary notion in the research community that social networks should be considered in forced migration models (Reinhardt et al., 2019; Hinsch & Bijak, 2019; Dekker et al., 2018; Simon, 2019; Al-Khulaidy & Swartz, 2020; Frydenlund & De Kock, 2020; Borkert, Fisher, & Yafi, 2018; Hilton & Bijak, 2017; Blumenstock, Chi, & Tan, 2019; Garip, 2008). The inclusion of these social networks at scale paves the way for further modeling efforts in forced migration studies that can further explore the opaque and vastly unknown interplay of social dynamics in refugee movement.

Contribution 2: methodological improvements on the state-of-the-art forced migration ABM to include publicly available computation tools and methods to facilitate the modeling community

The model presented herein both improves upon state-of-the-art FLEE model accuracy and replicates FLEE's computational speedup for modeling forced migration at scale. The largest methodological improvement that can be made to the FLEE model (and other computational models of similar caliber but less extensibility), is the automated creation of a simulation environment that creates an architecture through which refugees can be simulated to move at will rather than an architecture which permits them only to move to pre-determined locations such as urban centers or refugee camps. The creation of such a simulation environment in this model is not reliant on any pre-existing refugee infrastructure in an AOI, meaning that the model is applicable at the inception of a forced migration event. It is also therefore applicable in any geographic locality worldwide regardless of pre-existing refugee presence or infrastructure.

The computational models and methodologies developed as part of this research are made publicly available on GitHub at https://github.com/mrichey17/mig and the computational scalability aspects of this dissertation were presented via pre-recorded video at the Interservice/Industry Training, Simulation, and Education (I/ITSEC) conference held on 30 November – 4 December 2020 and are published in the conference proceedings (Richey & Mostowsky, 2020). ABMs and other sociocultural simulations are not trivial to develop for many reasons, one of which is timeliness (Frydenlund & De Kock, 2020). The research presented in this dissertation includes an automated methodology for the creation of a simulation environment referenced above which is publicly available on GitHub to jumpstart future modeling and simulation efforts for those who may wish to generate a simulation environment and use it to implement additional models.

1.5 Organization of Chapters

The research presented in the following chapters advances the art and the science of forced migration studies in the CSS domain such that the extant body of knowledge is now several steps closer to the authoritative decision support tools described above. Chapter 2 of this dissertation provides a review of relevant literature in the forced migration modeling domain, to include recent advances in computational modeling efforts, the need for and theory behind the inclusion of social networks in forced migration models, and the challenges of scalability in such models. Chapter 3 presents the methodology used to approach these challenges and provides a comprehensive Overview, Design concepts, and Details (ODD) document which communicates model instantiation and logic. Chapter 4 details the computational implementation of the model. Chapter 5 presents the results of the application of this methodology to a particular case study – Syrian migration into Turkey resulting from the ongoing Syrian Civil War – international conflict event. It also includes two additional case studies of Lebanon and Jordan to generalize the model across the whole of the Syrian Civil Conflict. Finally, Chapter 6 reiterates the impact this research has on the social science and CSS communities, and how it can be leveraged for further analytic inquiry in the discipline of migration studies moving forward.

2. BACKGROUND

This chapter provides a review of the relevant literature on migration that contributes to this research both theoretically and practically. Section 2.1 provides an overview of the taxonomy of migration literature reviewed in this dissertation. Section 2.2 follows with a discussion of relevant migration theory. Section 2.3 reviews extant migration modeling methodologies.

2.1 Introduction to Migration Literature

Migration research dates back to the late 19th century (Ravenstein, 1885). Over the years, migration research has addressed a variety of aspects of migration and made use of the techniques and methodologies available to researchers at the time of writing. Over the last two decades, the CSS community has brought more sophisticated computational modeling techniques to bear on the issue of migration and has begun to produce the types of models addressed in this dissertation. For the purposes of understanding a general taxonomy of migration studies through the years, Figure 2 presents a breakdown of the relevant areas of study to contextualize the research undertaken in Chapters 3 and 4.

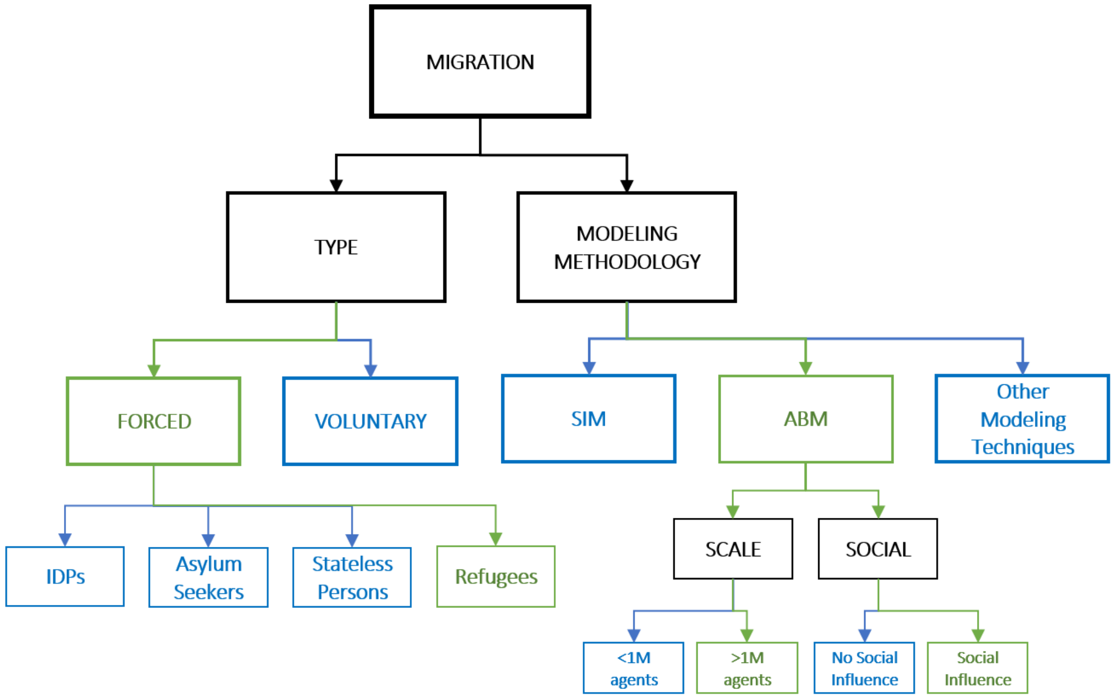


Figure 2 General taxonomy of migration studies highlighting the areas of contribution of this dissertation in green.

Taking into account the above taxonomy, the term migrant is interpreted to refer to anyone who moves from his or her location of origin to a new location. He or she can do so voluntarily, typically for reasons concerning livelihood or economic gain, or may be forced to move due to security, conflict, or environmental reasons (De Kock, 2019). If the migrant is forced to move, he or she may fall into the category of refugee, IDP, asylum seeker, or stateless person. The UNHCR defines a refugee as "someone who is unable or unwilling to return to their country of origin owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group, or political opinion" (UNHCR, 2020a). This definition is meant to contrast that of an IDP, who has not yet crossed an international border as a result of displacement. The factors that influence voluntary migrant decision-making and forced migrant decision-making must be kept separate regardless of modeling methodology (Adhikari, 2013; Richmond, 1993; Langley et al., 2016). In keeping, the primary type of migration is first distinguished as voluntary or forced. The topic of this dissertation is forced migration referred to in some literature as conflict-induced migration (Shackelford et al., 2020; De Kock, 2019; Echevarria & Gardeazabal, 2019).

The second component of the migration taxonomy is the modeling methodology, which is divided broadly into SIMs, ABMs, and other modeling techniques. This literature review will focus on SIMs, the modeling methodology that dominates the migration studies literature, and ABMs, the modeling methodology chosen for this research. That said, this is not intended to discount the numerous other modeling techniques that have been applied to migration studies over the years; namely, probabilistic or statistical modeling (Cohen et al., 2008; Bayar & Aral, 2019; Azose & Raftery, 2019; Bijak, 2006; Flowerdew & Lovett, 1988; Simini et al., 2012), cellular automata (Portugali, 1995; Dabbaghian, et al. 2010; Benito-Ostolaza et al., 2015), or microsimulation (Ballas, Clarke, & Wiemers, 2005; O'Donoghue, Cathal, & Lennon, 2010; Dekkers, 2015).

In Section 2.3 SIMs are reviewed and in Section 2.4, I turn the attention to ABMs. Within the taxonomy of ABMs, two distinct elements are addressed: scale and social networks. In the domain of forced migration, there are several ABMs but only one that provides the scalability required to model present day refugee crises (millions of agents), as outlined in Chapter 1. In terms of ABMs of migration, the few that address the social networking component are representing voluntary migration dynamics, and the two that address forced migration do so in a highly abstract, theoretical context without any empirical application or scalability. As the literature review dives more deeply into the taxonomy, it will arrive at the gap in the scientific community of migration studies that necessitates the research presented herein.

2.2 Migration Theory

Numerous behavioral and cognitive theories have influenced migration studies over the years, most notably the push/pull factor theory, the notion of herd behavior and network-assisted migration, the theory of destinations, and the theory of intervening opportunities. These theories are the product of sociological research into the factors that both influence and explain migration (Langley et al., 2016). These theories are influential in present-day modeling efforts as they reflect decades of sociological field and statistical research around why migrants migrate, how they decide where to go, the routes they ultimately devise, and the degree to which they interact with fellow migrants.

Push and Pull Factors

One of the most recognized theories pertaining to migration is the theory of push and pull factors that simultaneously repel and attract a migrant to a given location (Dorigo & Tobler, 1983; Langley et al., 2016). Lee (1966) once characterized spatial mobility as being influenced by factors associated with the origin, the destination, intervening obstacles, and personal factors, and this taxonomy became a cornerstone to many migration studies. The idea of push and pull factors makes a certain intuitive sense in the context of migration. Factors such as conflict, violence, threat of violence (Ibanez

& Velez, 2008; Shackleford et al., 2020; Bohnet, Cottier, & Hug, 2013; Naude, 2010), or loss of land (Adhikari, 2013) cause a person to consider moving, where factors such as greater economic prosperity (Durand & Massey, 2004; Massey et al., 1993) and a similar ethnolinguistic community in a destination (Barthel & Neumayer, 2015; Lin, Carley, & Cheng, 2016; Poot et al., 2016; Karemera, Oguledo, & Davis, 2000) might attract a person to a new area. The push and pull factor theory, however, is not without its criticisms with voices such as that of Castles (2003) observing that it fails to explain the null hypothesis, or lack of global migration given that the vast majority of people choose to remain in their countries of origin.

Herd Behavior and Network-Assisted Migration

The discussion of pull factors quickly leads to a discussion of herd behavior, or the tendency of migrants to go where others have gone before them (Epstein, 2008). Herd behavior is the early introduction in migration research to the inclusion of the social or network-based component within migration modeling (Epstein & Gang, 2006; Docquier, Peri, & Ruyssen, 2014; Arango 2000; Tranos, Gheasi, & Nijkamp, 2015; Havinga & Bocker, 1999; Robinson & Segrott, 2002). If people tend to go where others have gone, social networks and social media, then, are a migrant's glimpse into not only where other migrants have gone but where other migrants currently reside, and the social, political, and economic climate of the target location. In this sense, the idea that migrant decision-making should in some way incorporate social influence or social networks is not new (Klabunde & Willekens, 2016).

Dekker et al. (2018), Borkert, Fisher, and Yafi (2018), and Maitland and Xu (2015) make the case for the importance of social networks in papers detailing how Syrian refugees use smartphones to access social media and their social networks. Of the 54 refugees interviewed in Dekker's (2018) paper, smartphones were prevalent, and WhatsApp and Facebook groups were among the most prominent sources used to assist in the migration journey. Of the 150 Syrian refugees interviewed, WhatsApp groups were also among the most frequent means for identifying emergent and unofficial refugee populations in Turkey of similar ethnolinguistic background as the migrant – in the case of Syria, Kurdish or Arab (Richey, 2014a; Richey, 2014b). Dekker et al. (2018) note that two phenomena that result from a migrant using social networks to navigate a forced migration event are "the expansion of migration networks beyond existing ties and a diversification of available information on migration routes and destinations" (p. 2).

This recent finding demands, at minimum, the consideration of Social Network Theory (SNT) in forced migration modeling efforts. SNT addresses systems relationally, representing actors, agents, or in this case refugees, as nodes in a network, and the interactions and interrelations among them as edges or links (Borgatti & Ofem, 2010). While the migration literature emphasizes strong ties over Granovetter's (1977) seminal weak ties, it is precisely these weak ties recent research is showing plays a role in forced migration decision-making. The creation of new ties that Dekker et al. (2018) noted may very well be weak ties but may also be highly impactful to refugee decision-making in reactive, forced, and conflict-induced migrations scenarios.

Theory of Intervening Opportunities

The theory of intervening opportunities is prominent in the migration literature and even made its way into Lee's (1966) seminal characterization of migrant decision-making. It states that a migrant's decision where to migrate is proportional to the opportunities that exist in that candidate location and inversely proportional to the hazards of the journey or other candidate destinations that may present themselves along the way (Stouffer, 1940; Stouffer, 1960). In short, this theory suggests that both the security and opportunity afforded a refugee at the destination location is more important in the refugee's decision-making than the hazards of the journey itself.

Theory of Competing Destinations

The theory of competing destinations was put forth largely to nuance the prevalence of the gravity model in migration studies (explained more deeply in Section 2.3). This theory holds that the likelihood of a migrant moving to any one location cannot be calculated without assessing the likelihood of moving to all other possible locations (Fotheringham, 1983). Present-day models that include the theory of competing destinations should, then, incorporate a parameter or some other function that takes into account all possible destinations or, at minimum, an opportunity cost for choosing one destination over another.

2.3 Migration Modeling Methodologies

For simplicity and expediency, this review of migration models will address two primary categories of computational models: SIMs and ABMs. SIMs and seminal applications of the gravity model along with more nuanced competing destinations models, comprise the bulk of the migration literature, particularly voluntary migration. There are only a handful of ABMs in the literature and fewer that address forced migration, but they are reviewed here with their relation to the topic of this dissertation.

Spatial Interaction

Traditional methods for modeling and, to a lesser extent, forecasting mass migration have made substantial use of the gravity model and SIMs (Goodchild & Smith, 1980; LeSage & Fischer, 2010; Rae, 2009; Rogers et al., 2002; Sarra & Signore, 2010; Stillwell, 1978; Vernon-Bido et al., 2017). These models make use of international in- and out-migration statistics over the course of decades to quantify bilateral flows of migrants between locations. Perhaps one of the most well-grounded network-based SIMs is Simpop, which represents urban areas as multi-agent systems with nodes (cities) and edges (interactions between urban centers, e.g., the movement of people) (Pumain et al., 1995; Bura et al., 1996; Bretagnolle & Pumain, 2010). Other more preliminary forms of spatial interaction are purely statistical, such as the fitting of regression models to bilateral migration flows (Flowerdew & Lovett 1988) or implementations of the power law, or Zipf's law as in Simini et al. (2012).

Initially conceived to explain physical movement, the gravity model was subsequently adapted to explain phenomena occurring in international trade (Matyas, 1997; Bergstrand, 1985), linguistic diffusion (Trudgill, 1974, Nerbonne, 2010), and, eventually, migration (Karemera et al., 2000; Lewer & Van Den Berg, 2008; Lin, Carley, & Cheng, 2016). The gravity model stipulates that migration flows are related to the population sizes of both origin and destination locations, and, secondarily, the geographic distances between them (Sarra & Signore, 2010). Simply put, the original gravity model is loosely specified in Equation (1) in which M_{ij} is the interaction between location i and location j, P_i and P_j is the population of location i and location j, and d_{ij} is the distance between the two locations.

Equation 1 Gravity Model

$$Mij = \frac{PiPj}{(dij)^2}$$

The gravity model's application to the field of migration studies and its evolution to more robust SIMs quickly led to a family of SIMs. These models were then applied to modeling voluntary migration as a spatial and econometric problem. Examples of these applications include those found in LeSage and Fischer (2008), and Poot et al. (2016).

The group of SIM models includes: 1) the unconstrained model, 2) the production-constrained model, 3) the attraction-constrained model, 4) the production-attraction-constrained model, and variations on these themes (Wilson, 1970). Briefly, the unconstrained model assumes transitivity in bilateral interaction and is specified within the original conceptualization of the econometric gravity model. The production-constrained model assumes a known quantity at origin and an unknown distribution pattern. Such a model could be used to estimate retail sales in shopping centers or may very well apply to the problem of forced migration where there is a known quantity of refugee migrants at a point of origin such as a border crossing with unknown destinations in the target country. The attraction-constrained model assumes that destinations are known, such as workplaces, and origins are yet-to-be-determined, such as dwellings. The production-attraction-constrained model is largely employed in transportation studies and focuses on estimating the interaction M_{ij} given input on origin and destination locations (Wilson, 1970). Largely used to estimate interzonal migration (Stillwell, 1978; Fotheringham & Kelly, 1988), and at times international migration (Yano et al., 2000, Yano et al., 2003), but always voluntary migration, production-constrained models have dominated the migration literature. Sarra and Signore (2010), for example, implement an origin-constrained model in Poland capable of measuring the magnitude of pull factors at different time steps.

As indicated in Section 2.2, competing destinations models were developed in response to and in critique of the gravity model. With this model, Fotheringham (1989 & 1983) asserts that "gravity models are misspecified since they do not include a variable which explicitly measures the relationship between interaction and competition between destinations" (p. 21) By adding a variable representative of competition between destinations (much as Trudgill (1974) added a variable to the gravity model representative of linguistic similarity), a new set of spatial interaction models were produced. These models were extended further still to incorporate various methods, such as eigenvector spatial filtering, to account for the spatial and network autocorrelation potential Fotheringham initially uncovered (Chun & Griffith, 2011). Finally, cell-space or cellular automata models were attempted as a means to contextualize migration within city planning, but these attempts were much less focused on the process of migration itself as the subject of study (Portugali et al., 1995). While the idea of herd behavior (reviewed in Section 2.2) introduced the idea of social networks conceptually to migration studies, gravity models levied the structural theory that migration flows were not only spatial in nature, but network based. This led to a vein of research that applied network scientific methodologies to the understanding of migration.

As both SIMs and Geographic Information Systems (GIS) matured in the late 80s and 90s, the ability to visualize migration (and spatial movement in general) using flow mapping techniques became more mainstream (Tobler, 1987 & Rae, 2009). During this time, “the convergence of data manipulation, visualization, and spatial analysis” gave rise to a new wave of spatially explicit simulations (Benenson & Torrens, 2004). These simulations included examples like modeling migration as the path of least resistance across a spatially explicit landscape filled with barriers to movement (Cushman et al., 2006, Cushman et al., 2010; Aral et al., 2009). This slightly more robust take on an origin-constrained model marked the beginning of more advanced simulation efforts. Despite these efforts, SIMs of migration are insufficient for two primary reasons.

First, traditional methods focus on voluntary versus forced migration. While voluntary migration is important, the methods and techniques used to model voluntary versus forced decision processes are not the same. Traditional voluntary migration models tend to be econometric in nature weighing economic factors more highly than other factors in the model if other factors are considered at all. Second, SIMs lack the granularity and control for modeling migrant decision-making at the individual level. The decision to migrate is a personal one undertaken either as an individual or as a family unit. As detailed in Section 2.2, increasingly migrants are using social media to monitor their social networks to identify candidate destinations identified by family members, acquaintances, or other refugees. SIMs do not account for the increasing importance of social media and social networks in migration modeling, nor do they provide a computational environment where individual migrants can make non-deterministic decisions about the migration process. Most significantly, however, SIMs and system dynamics approaches are part of a macro-level, top-down conceptualization of systems and subsystems (Batty, 2008). Over the past several decades in CSS, our collective understanding of complexity in the social sciences and how to represent that complexity computationally has evolved tremendously (Miller & Page, 2009; Gilbert & Troitzsch, 2005; Simon, 1996). SIMs, at least for the analytical inquiry surrounding forced migration, do not satisfy the desire for bottom-up understanding of sociocultural phenomena and the instantiation of environments from which complex social patterns and systems can emerge. Gray, Hilton, and Bijak (2017) call for three deeply considered constructs in agent-based spatial demography: time, uncertainty, and heterogeneity. SIMs cannot support at least two of the modeling constructs in this call. For this ability, we turn to ABMs.

Agent-Based Models

Given that migration, whether forced or voluntary, is a very personal decision taken either as an individual or as a family unit, ABMs are an ideal modeling technique for this particular sociocultural phenomenon. By allowing individual decision-making within a simulation environment, ABMs provide a construct for the emergence of non-deterministic spatial and social patterns among refugee agents. ABMs provide a natural approach in the computational analysis of forced migration because they allow the simulation of behaviors of heterogenous agents while observing resultant system evolution over time (Crooks et al., 2008). The ability to encode decision-making at the individual level into a model of forced migration is both important and possible because migrant decision-making is not random (Kennedy, 2012), though, as has been shown in recent research, it may, due to social influence, be sub-optimal (Simon, 2019; Hinsch & Bijak, 2019), discussed more thoroughly in Section 2.4. Another strength of ABMs is that they can be instantiated and tested with a variety of topologies; for example, network, spatially explicit or geospatial, Euclidian space, grid-cell (cellular automata), etc. (De Kock, 2019). When it comes to shorter term, though sometimes larger scale, reactive or forced migration events however, it is the general patterns of migration ABMs will excel at predicting, not the precise locations of specific refugee outflow (Edwards, 2008).

There are many small-scale, one-off models have been applied to forced migration contexts; for example, Hebert, Perez, and Harati (2018); Collins and Frydenlund (2016); Hattle (2016); Frydenlund et al. (2018); Sokolowski, Banks, and Hayes 2014; and Sokowlowski et al. 2014). The FLEE model is the most robust of recently developed ABMs to model forced migration and the only one that approximates any kind of generalizable modeling framework (Suleimenova et al., 2017; Suleimenova & Groen, 2019; Suleimenova & Groen, 2020). The FLEE model has been run and tested empirically across five independent geographic areas most within the African continent and the Python 3 codebase is designed for parallel computing architecture (Groen, 2018; Groen, 2019). It is not, however, without its considerable limitations. First and most critically, the instantiation of the simulation environment relies on pre-existing official refugee camps. This limits the FLEE’s applicability to forced migration events in two very important ways. Quite simply, the FLEE model cannot be applied at the beginning of a refugee crisis or conflict event before aid organizations have had the opportunity to establish any official temporary accommodation architecture in the region. This renders the model useless for predicting forced migration flows until a refugee crisis is well underway. The second limitation to the FLEE model is in spatial extent in that the model can only be applied to regions where refugee camps are present as these camps comprise the nodes in the network-based simulation architecture. This precludes FLEE from being applied in regions such as Eastern Europe where there is no official camp infrastructure. Additionally, in AOIs like Turkey, official camp infrastructure exists in a very limited spatial extent relative to the entire country although refugees are found across the nation. In this case, FLEE would only be able to model a fraction of the population – not the most important fraction given that the UNHCR reports that 98 percent of Turkey's refugee population resides in urban or rural areas outside of official camps (UNHCR, 2020b). Both of these limitations, in temporality and spatial extent, are derived from the methodology used to create the simulation environment. The creation of the simulation environment is additionally a highly manual process that requires many man-hours of effort to create even at constrained spatial extents. This manual data entry affects the timeliness of simulations, one of the biggest drawbacks to creating ABMs in response to crisis situations noted by Frydenlund and De Kock (2020).

It has been noted previously that one of the primary limitations of ABMs in general is the lack of verification and validation methods that exist for simulations (De Kock, 2019). In the case of FLEE, another flaw in the simulation design is the fact that the final refugee population in the simulation used for validation purposes of the model is also used as model input and as data within the model's subsequent steps. This makes the model somewhat of a self-fulfilling prophecy insofar as the modeler prescribes the number of agents at model initiation, the number of agents at model termination, and the simulation then predicts where those agents are likely to go within a network infrastructure of official refugee camps. In all these ways, the FLEE model lacks granularity, practicality, applicability, and realism for modeling forced migration, though it is, as previously mentioned, the closest thing the CSS and migration studies communities have to a generalizable framework for forced migration modeling.

Clearly, gaps remain in the application of scalable, realistic ABMs to forced migration contexts. While the body of work is expanding rapidly and this is very encouraging to the migration studies community, "without more investment from social scientists who study forced migration from a wide variety of perspectives, lenses, and methodological approaches, we have yet to tap the full power of simulation to advance forced migration theory, practice, and policy" (Frydenlund & De Kock, 2020, p. 63). The application of ABMs in this dissertation is one such lens.

2.4 Background on the Consideration of Social Networks in Modeling

When it comes to modeling forced migration populations, the consideration of social networks is scarce in recent scholarship. Recent theoretical research at the nexus of ABMs and forced migration suggests implications for how social networks are to be implemented in model design. For example, a notion featured prominently in ABM design and development is the idea of path-dependent variation, or the idea that multiple paths all resulting from plausible causation must lead to the sociocultural event, pattern, or behavior in the context of simulation (Cioffi-Revilla, 2005). In this notion, Cioffi-Revilla refers to conceptual paths, but in the context of forced migration, we can interpret it to mean literal paths and migration routes. Simon (2019) set out to create a small-scale ABM of Mexican migration that included policy implications and migration routes. She found that while SNT predicts path-dependence, new migration pathways still emerge despite the robust inclusion of social networking elements. Her findings concluded that once a critical mass of migrants has settled on a destination and method of arrival, migration corridors will be robust to policy fluctuations and other exogenous factors (Simon, 2019; De Haas, 2010; Massey et al., 1993).

Recent theoretical research demonstrates how refugees using social networks and social ties in agent-based simulations may make less-than-optimal decisions and deviate from rational choice (Simon, 2019; Reinhardt et al., 2019; Hinsch & Bijak, 2019). This may explain why rational choice or utility/reward maximization theories are likely not the best decision theories to apply in forced migration contexts, unless the utility function includes a social element. Reinhardt et al. (2019), for example, demonstrated that migration routes are self-organized and predicated on social contracts. The highly abstracted and theoretical migration model in question was implemented simultaneously in two separate programming languages (Julia and ML3) with drastically different simulated results, yet both sets of results demonstrated migrants' reliance on social contracts to make transit decisions. Hinsch & Bijak, (2019), in an extension of the abstracted model above, found that migration routes "are an emergent property of the interactions between individuals" (p. 5), and that the more communication that exists between migrants, the more migrants concentrated on a few primary migration routes, and the less optimal those migration routes were. All of this work suggests that migrants are prone to use self-organized migration routes even if suboptimal when using their own social networks to determine where, how, and when to move. This concept is not unique to migration studies but is rooted in sociological analysis of social influence on group dynamics. Centola and Macy (2007) found that if agents only consider the actions of immediate neighbors, less-than-ideal norms can emerge and spread rapidly. In this way, it may be instructive for forced migrant agents to be more naïve than savvy, for example, operating with a limited or incomplete understanding of the surrounding simulation environment (DeAngelis & Diaz, 2019) and leveraging information passed through the weak ties social media and social networking tends to promote.

Many studies have additionally addressed social network mechanics that are used either for information transmission (Klabunde, 2014; Barbosa Filho et al., 2011; Rehm 2012 & Biondo et al., 2013; Al-Khulaidy & Swartz, 2020; De Haas, 2010) or for a more abstract exchange of social capital (Garcia-Diaz & Moreno-Monroy, 2012; Reichlova, 2005; Massey & Senteno, 1999). All of this research, however, provides examples of social networks in models of voluntary migration. None of these social concepts have been applied to models of forced migration or to refugee decision-making until quite recently. Three recent theoretical ABMs explore the implications of social networks and social influence in the context of forced migration route formation (Reinhardt et al., 2019; Hinsch & Bijak, 2019; Simon, 2019).

These recent findings demand, at minimum, the consideration of Social Network Theory (SNT) in forced migration modeling efforts. SNT addresses systems relationally, representing actors, agents, or in this case refugees, as nodes in a network, and the interactions and interrelations among them as edges or links (Borgatti & Ofem 2010). While the migration literature emphasizes strong ties over Granovetter's (1977) seminal work of the strength of weak ties, it is precisely these weak ties recent research is showing plays a role in forced migration decision-making. The creation of new ties that Dekker et al. (2018) noted may very well be weak ties but may also be highly impactful to refugee decision-making in reactive, forced, and conflict-induced migrations scenarios.

2.5 Summary of Findings

In summary, the literature has shown that SIMs are not the best choice for modeling forced migration because it is a top-down modeling approach that does not allow for the capture of individual migrant decision-making and the emergence of migration patterns in aggregate. ABMs present a viable bottom-up modeling approach but the majority of ABMs developed within the migration studies literature have been applied to voluntary migration contexts. Of those that have been applied to forced migration contexts, only one is empirically validated and designed to scale to millions of migrant agents – the FLEE model. This singular ABM of forced migration does not include a social network component, which the literature has shown should be considered in forced migration modeling paradigms. Social networks have been considered in both models of voluntary migration and models of forced migration, the latter only in a theoretical and highly abstracted context. This then presents a gap in recent scholarship that can be filled with an empirical and scalable ABM of forced migration which considers social networks. This dissertation constitutes this first empirical attempt and makes use of the theories around agent sensing and agent communication the theoretical considerations of SNT suggest, covered in more depth in Chapter 3, Section 3.2.

3. METHODOLOGY AND ODD

This chapter provides details on the methodology used to approach the research objectives introduced in Chapter 1. Section 3.1 addresses the general methodology while Section 3.2 presents a comprehensive Overview, Design concepts, and Details document (Grimm et al., 2020) report for model design and implementation of the model and model logic.

3.1 Methodology

The methodology for approaching the three research questions and objective is outlined in Chapter 1 and available in Figure 1. It is based loosely on the methodology for social simulation presented in Gilbert and Troitzsch (2005, p. 18-25). Here the methodology departs from Gilbert and Troitzsch (2005) by conflating the conceptual design and development phase. It consists of four main phases: Define, Review, Develop, and Apply. A further departure is to include model verification as subordinate to the Develop phase and model validation as subordinate to the Apply phase.

In the Define phase, the research objective is identified. In Chapter 1, the research objective was stated as developing an empirical computational model that advances our understanding of forced migration and considers social networks. open-sourceIn Chapter 2, it was outlined that the computational model should be an ABM. Further, that this ABM would fill a gap in extant scholarly research by providing the first scalable computational model of forced migration to consider social networks. In the Review phase, the relevant literature is surveyed and techniques, design concepts, and computational strategies are noted for possible inclusion within the model. This literature review is included in Chapter 2 and contains model design elements from forced migration theory such as the design of agent decision-making and awareness of the simulation environment. In the Develop phase, an ABM is first designed conceptually and then developed computationally. Verification is achieved by employing code reviews. RQ1 resides within the Develop phase in that a computational methodology is designed and developed to contribute to the body of social science knowledge. The final phase is Apply where the model is calibrated, applied to, and validated against a relevant case study – in this case, the Syria/Turkey, Syria/Lebanon, and Syria/Jordan forced migration events. Computational tools and techniques that can be repurposed for further modeling efforts are extracted from the Apply phase in satisfaction of RQ2. RQ3 entails the application of the model to the case study for validation purposes.

3.2 ODD of Agent Behavior and Model Logic

For the purposes of describing model functionality, model logic, and agent decision-making, an ODD protocol is provided below in addition to the Model Logic diagram found in Figure 3. The functionality of the model is important to understand such that the methods and results can be replicated across the research domain and in other research contexts. For this replication to be successful, a paper needs to describe not only the emergent behavioral patterns and results of the ABM, but also describe the design constructs from which that behavior arises. (Grimm et al., 2020). What follows is an in-depth description of the functionality and implementation of the ABM of forced migration described in this dissertation.

Overview

Purpose

The overall purpose of the model is to model and predict forced migration patterns considering the social context of the migration population within a given spatial extent during and directly following a forced migration event. The case study in question for this particular instantiation of the model is Syrian refugees migrating into and within Turkey. The model provides the research community with a tool that facilitates better understanding of forced migration and addresses all three RQs of this dissertation. The intended predictive window for the model is 30-90 days. The model described in detail below is referred to as the 'full' model, or Condition 1, throughout the remainder of the dissertation.

Entities, State Variables, and Scale

The model includes the following entities:

- Refugee agents
- Location nodes

Refugee agent is a single agent class in the model that represents refugees, asylum-seekers, and stateless persons. They are naïve agents and, as such, do not have many attributes except their locations and their social networks, which are specified randomly using friendship and kinship ties as described in Table 1.

Table 1 State variables of each simulation agent.

Variable Name	Variable Type & Range	Variable Meaning
num_friends	Dynamic integer range, 0…*n* Suggested range: 0-5	The number of friendship ties a migrant agent has in its social network
num_kin	Dynamic integer range, 0…*n* Suggested range: 0-5	The number of kinship ties a migrant has in its social network
Location Latitude (not an editable parameter)	Dynamic float Determined by simulation	The latitude of the location node where the migrant agent is located at any given time step *t*
Location Longitude (not an editable parameter)	Dynamic float Determined by simulation	The longitude of the location node where the migrant agent is located at any given time step *t*

The number of friends and kin in the simulation are the simulation parameters that correspond to the implementation of social networks. Social ties are created randomly in the simulation design and the treatment of friendship ties and kinship ties is the same, though they can be weighted differently at such time as parameter weights are set

Location nodes. Location nodes are points within a larger undirected geographical network comprising the simulation environment. Location nodes are created with the desired granularity of the researcher using the geographical centroids of typically the first, second, or third level administrative boundaries for the target AOI. For a country such as Turkey, it is recommended that either the first or second administrative levels be used, resulting in a location network of 81 nodes or 929 nodes, respectively. Location nodes are also connected using bidirectional links to every other location node within the network with which the target location node shares a geographic border. Some location nodes are additionally specified as border crossings, which represent nodes in the geographical network where migrant agents can enter the simulation by crossing an international border. In the broader simulation architecture, agents can be seeded anywhere in the network, not just at border crossings. Additionally, nodes are characterized by hosting a violent event (conflict), an official refugee camp (asylum), both, or neither.

The rationale for using the geographic centroids of administrative areas to create the geographical network that comprises the simulation environment is to provide the most organically derived movement options for forced migrant agents to transit terrain. Previous simulations have derived the simulation environment exclusively from pre-existing refugee camp locations (Suleimenova et al., 2017), from urban centers (Pumain et al., 1995; Bura et al., 1996; Bretagnolle & Pumain, 2010), and from a grid-cell matrix overlay on the geographic terrain (Cushman & Chase, 2015). The use of pre-existing refugee camp locations as node locations within the network is necessary but not sufficient for realistic modeling. For instance, the UNHCR reported that 98 percent of Turkey's refugee population resides in urban or rural areas outside of Temporary Accommodation Centers, or camps (UNHCR, 2020). If the majority of a country's refugee population resides outside of official camps, and camps open and close over time, then the exclusive reliance on pre-existing official refugee camps as the candidate locations for migration in a simulation is not realistic as seen in the FLEE model (Suleimenova et al., 2017). The simulation's observer state variables appear in Table 2.

Table 2 Observer state variables of the simulation environment.

Variable Name	Variable Type & Range	Variable Meaning
camp_move_probability	Static float, 0-1 Suggested range: 0.1-0.5	Percent probability that a refugee agent will decide to move from its current location node if a refugee camp exists at that location node
conflict_move_probability	Static float, 0-1 Suggested range: 0.5-1	Percent probability that a refugee agent will decide to move from its current location node if a conflict event occurs at that location node
other_move_probability	Static float, 0-1 Suggested range: 0.7-1	Percent probability that a refugee agent will decide to move from its current location node
seed_refs_per_node	Static integer, 0…*n*, or dynamic integer range Suggested range: dependent on real-world conditions	Number of refugees that enter the simulation through a border crossing at each time step
seed_nodes	Static list of integers Suggested input: dependent on real-world conditions	A list of node locations designated as border crossings where migrant agents enter the simulation at each time step
anchor_location	Static float Suggested input: dependent on real-world conditions	Latitude and Longitude of a general geophysical anchor point beyond the spatial extent of the simulation towards which the majority of migrant agents are moving
num_steps	Static Integer Suggested range: 30-90	Number of desired simulated days. Each step is equivalent to one calendar day

In addition to the observer state variables of the simulation, two variables affect the computational implementation of the simulation. These two variables described in Table 3 are editable and allow the simulation to be modified to fit the hardware constraints of the virtual or physical computational environment in which the simulation is to be run.

Table 3 Variables affecting implementation of the simulation on virtual or physical hardware.

Variable Name	Variable Type & Range	Variable Meaning
num_batches	Static integer Suggested range: 1-16	Number of batches of migrant agents to create for parallelization of the model. The number of batches of agents does not need to correspond to the number of CPU cores available for multiprocessing.
num_processes	Static integer Suggested range: 1-16	Number of parallel processes provisioned for the movement of migrant agents. Notionally, the number of processes should correspond to the number of CPU cores on the provisioned virtual or physical simulation machine

The first three observer state variables and the anchor_location are probabilistic mechanisms in the simulation for determining if and when agents will move decision-making. Thresholds for these parameters are explored through sensitivity analysis, per Gilbert and Troitzsch's (2005, p.24) design for social simulation. The remaining variables allow the researcher to adjust aspects of the simulation environment, to include the locations at which new refugee agents enter the simulation, the number of new refugee agents that enter the simulation at each time step, and the number of batches into which refugee agents are split for multiprocessing. The first two of these final three variables (seed_nodes and seed_refs) should be determined during calibration of the simulation to real-world conditions during the target time period through contextual research of the forced migration crisis in question. Location nodes are characterized by two endogenously calculated parameters that forced migrant agents calculate for each candidate move location once the decision has been made to migrate. The first dynamically calculated parameter is the node's location score and the second is its desirability score. Each score is calculated differently for each candidate location node by each migrant agent at each time step, so no single node will look the same from the vantage point of different agents. This is discussed further in the ODD section on stochasticity and an example of this calculation is provided in the following section. The presence of conflict events is presented as a push or deterring factor within the calculation of the desirability score (Langley et al., 2016; Ibanez & Velez, 2008; Naude, 2010a & b). The presence of refugee camps is presented as a pull factor within the calculation of the desirability score. Factors are aggregated and the location node with the highest desirability score in the migrant's range is the node to which the migrant agent moves at the given time step. The degree to which factors contribute to a location node's desirability score is presented as editable parameter weights in Table 4 and detailed explanations of these calculations can be found in the following section.

Table 4 Editable parameter weights to weight the influence each variable has on a location's attractiveness to refugees.

Variable Name	Variable Type & Range	Variable Meaning
population_weight	Static integer range, 0-1 Suggested range: 0.5-1	Normalized between 0 and 1 number of migrants at a given location node at each time step
location_weight	Static integer range, 0-1 Suggested range: 0.5-1	Normalized between 0 and 1 indexical score of how close the node location is to the Location variable relative to every other location node in the network
kin_weight	Static integer range, 0-1 Suggested range: 0-0.3	Normalized count * variable weight
friend_weight	Static integer range, 0-1 Suggested range: 0-0.3	Normalized count * variable weight
conflict_weight	Static integer range, 0-1 Suggested range: 0.3-0.7	Boolean presence of conflict * variable weight

Variable Name	Variable Type & Range	Variable Meaning
camp_weight	Static integer range, 0-1 Suggested range: 0-0.5	Boolean presence of camp * variable weight

The spatial extent of this model is administrative level 0 of the country of Turkey encompassing approximately 300,000 square miles of geographic terrain. Though Syrian migrants are moving from Syria into all neighboring countries to include Jordan, Lebanon, Iraq, and Iran, over half of Syria's refugees are migrating to Turkey with the final destination of Istanbul or on into Eastern or Central Europe (Icduygu, 2015). This model is focused on the Syria-Turkey migration use case and, as such, is focused on refugees and asylum seekers as opposed to IDPs. The model is currently configured using the spatial resolution of administrative level 2 which results in 929 individual district-level location nodes in the network. The model could just as easily be configured to administrative level 1 with 81 provincial-level location nodes. In keeping with the theory of intervening opportunities (Stouffer, 1940; Stouffer, 1960), distance across or between location nodes (and therefore the time required to transit them), is not explicitly accounted for within the simulation. Additionally, the simulation environment is so granular that the largest transit distance between two nodes would not result in travel in excess of one day. The temporal resolution of this model is indefinite. One model time step t represents one day and the nature of model predictions are intended for 30-90 day future predictions of migration flows.

Process, Overview, and Scheduling

The model is variable-based (Quigley & Marina, 1979) versus event-based (Cioffi-Revilla, 2005) in that agents are activated randomly at each time step and make decisions based on deterministic and stochastic variables, not catalyst events that occur at certain time steps. Forced migrant agents also do not directly interact in the simulation; rather, interact indirectly via the simulation environment and their social networks. This mediated interaction is ideal for computational reasons insofar as it eliminates the need for direct computational agent-to-agent interaction at every time step. Agents store the information they require from other agents in their social networks as indices and receive updates to these indices at every time step. Direct interactions would make read/write operations to the simulation environment and agent decision-making highly computationally complex.

The model parallelization and logic processes that repeat each time step are:

- **Calculate**: Calculate or re-calculate deterministic portion of node location score
- **Split**: Calculate indices to batch migrant agents and into *x* number of groups for decision-making parallelization
- **Move**: Execute agent decision-making and movement. For each migrant agent in group:
    - **Activate**: Decide whether or not to move from current location in accordance with percent move chance probabilities.

    - **Gather**: Fetch neighboring location list as a list of move location candidates
    - **Assess**: Compute desirability scores for each candidate location based on environmental (location) and social (agent) state variables
    - **Move**: Create new agent at location with highest desirability score.
    - **Return**: From each process, return a list of new refugee objects and a dictionary to store the refugees at each node after moving
- **Update**: Aggregate refugees from each process and overwrite refugee objects from previous timestep. Aggregate dictionaries of refugee locations. Calculate new weights from aggregated dictionaries and updated location entities in the graph.
- **Create**: Create relationships between co-located refugees.

These processes are represented visually in Figure 3.

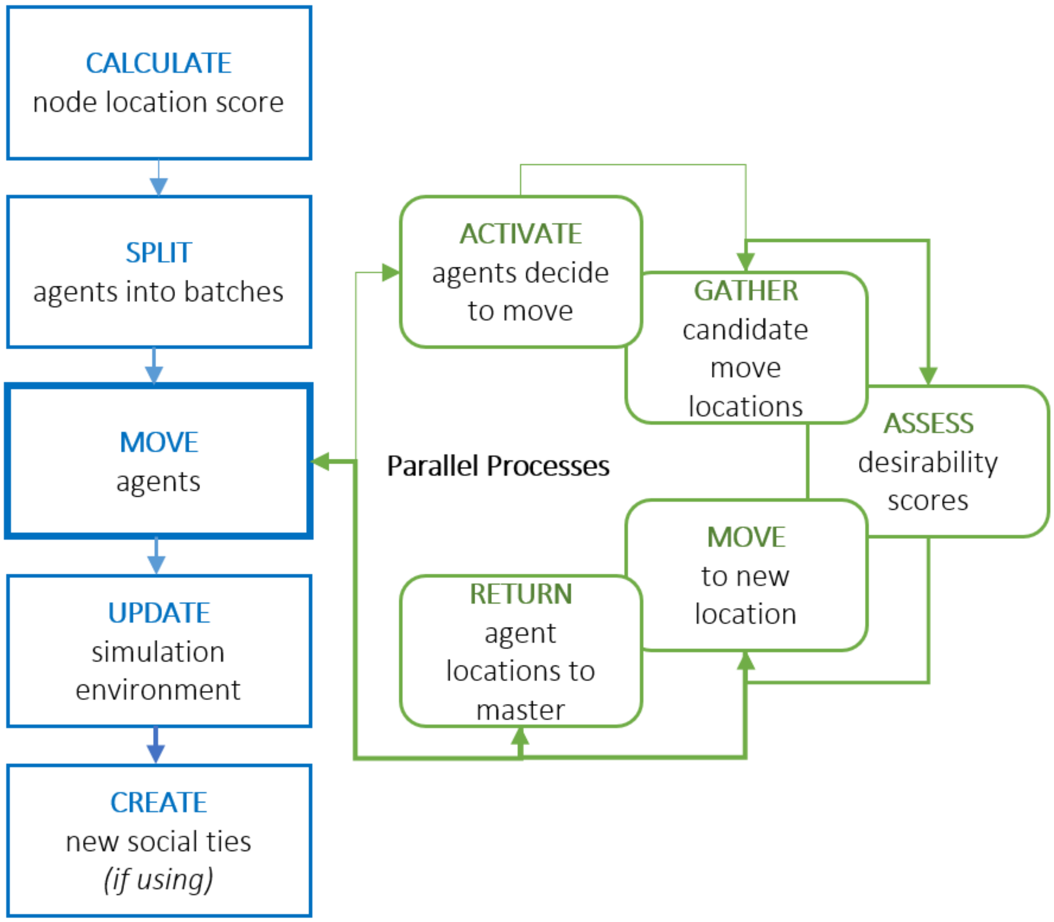


Figure 3 Model logic overview and flow diagram depicting what occurs at each model step, inclusive of the actions that occur in parallel across multiple batches of agents.

In the first parent process, Calculate, node location scores are determined for every location node in the simulation environment. This calculation is made as follows:

Equation 2 Node location calculation

$$\mathrm{E} = (p \times w_1) + (l \times w_2) + (c \times w_3) + (d \times w_4)$$

where p is the normalized existing refugee population in the candidate node location, l is the normalized proximity of the candidate node location to the general destination location, c is the presence of official refugee camps in the candidate node location, d is the presence of security-related or conflict events in the candidate node location, and w_1 through w_4 are the respective variable weights as specified by the modeler at model initiation.

In the second parent process, Split, agents are batched for parallel processing using the variables num_batches and num_processes. While the number of parallel processes should generally correspond to the number of CPU cores on the physical or virtual machine available for computation, the number of agent batches can equal or exceed this value. For example, on a physical machine with 8 CPU cores, an ideal number of parallel processes could be 6, reserving two CPU cores for other ongoing computer operations. The number of agent batches, however, could be anywhere from 6 to 10. This is because batches of agents, while containing the same number of agents, will not necessarily require the same amount of parallel processing time. The size of an agent's social network is one of the key drivers of latency in model calculations and, as such, agents with larger social networks (10-15 social ties) will take more time to execute decision-making logic than agents with smaller social networks (1-5 social ties). For this reason, two batches of 500,000 agents may yield different overall processing times. Having more agent batches than there are available CPU cores ensures that a given core, once execution is complete across its assigned batch of agents, can begin processing decision-making logic for a second batch of agents while other more complex batches of agents continue to be processed. The program will execute this selection automatically if the number of agent batches is higher than the number of parallel processes. The parallel processing model mechanics are only available in a Linux environment due to the challenges associated with multiprocessing in Python.

So far, all parent processes have been executed serially. In the third parent process, Move, parallel processing begins. Parallelization is achieved using multiprocessing techniques in Python 3, meaning that each parallel process maintains a separate or distributed memory environment. The simulation environment is copied over to each child process at the beginning of Move so a copy of the same simulation environment is available to each child process. Batches of agents, then, move to their assigned core to execute their model logic, bringing with them their geolocations and information pertaining to their social ties stored as indices.

Agent decision-making logic is executed in parallel across the five child processes: Activate, Gather, Assess, Move, and Return. During Activate, agents follow a probabilistic activation, drawing from a normal distribution, in deciding whether or not to move from their origin location given the user-specified parameters conflict_move_probability, camp_move_probability, and other_move_probability. If an agent decides to move, the agent progresses to Gather, where it collects the node location scores for all candidate move locations one degree of separation away from the agent's origin location node. In the child process Assess, agents then calculate and add the stochastic elements of their decision-making logic to the node location scores calculated previously. The result of this calculation, made by each agent for each candidate location node at each time step in parallel, is known as the desirability score, and is calculated as follows:

Equation 3 Node desirability score calculation

$$\mathrm{N} = \left((k \times w_5) + (f \times w_6)\right) + \mathrm{E}$$

where k is the normalized number of kinship ties belonging to a migrant agent in the candidate node location, f is the normalized number of friendship ties in the candidate node location, and w_5 and w_6 are the respective variable weights as specified by the modeler at model initiation. E is the output of Equation 2.

For clarity, an example calculation is provided in Table 5.

Table 5 Example calculation.

	Location 1	Location 2	Move Decision
Agent 1	E = 0.45 N = ((0.1*0.5) + (0.9*0.5)) + 0.45 N = 0.95	E = 0.25 N = ((0*0.5) + (0.5*0.5)) + 0.25 N = 0.5	Location 1
Agent 2	E = 0.45 N = ((0.3*0.5) + (0*0.5)) + 0.45 N = 0.6	E = 0.25 N = ((0.8*0.5) + (0.7*0.5)) + 0.25 N = 1	Location 2

Two agents at the same origin location assess the same two move location candidates and reach different decisions. The deterministic elements will ensure that the result of the first equation is the same for both agents, assumed E = 0.45 for Location 1 and E = 0.25 for Location 2, calculated per Equation 2. Agent 1 has friends and some family in Location 1 and some friends in Location 2. Agent 2 has many friends and family in Location 2 and some family in Location 1. For simplicity, all parameter weights are assumed to be equal at 0.5.

In the child process Move, the agent then moves to the determined location, and in the child process Return, it reports its new location back to the master simulation environment while simultaneously removing itself from its previous location. Parallel processing ends here. The child processes are represented visually in more detail in Figure 4. Equation 2 is the calculation that produces E in Table 5 and Equation 3 is the calculation that produces N in Table 5. Figure 4 depicts what occurs within the parallelized child processes and how agents reach move decisions. Gray boxes represent editable parameters in the simulation (move, friends, kin), variables derived from input data (proximity, camps, conflict), or derived variables calculated within the simulation (refugee population). Black arrow strength represents editable probability distributions in the Activate process. Black arrow strength represents different editable parameter weights in the Assess process.

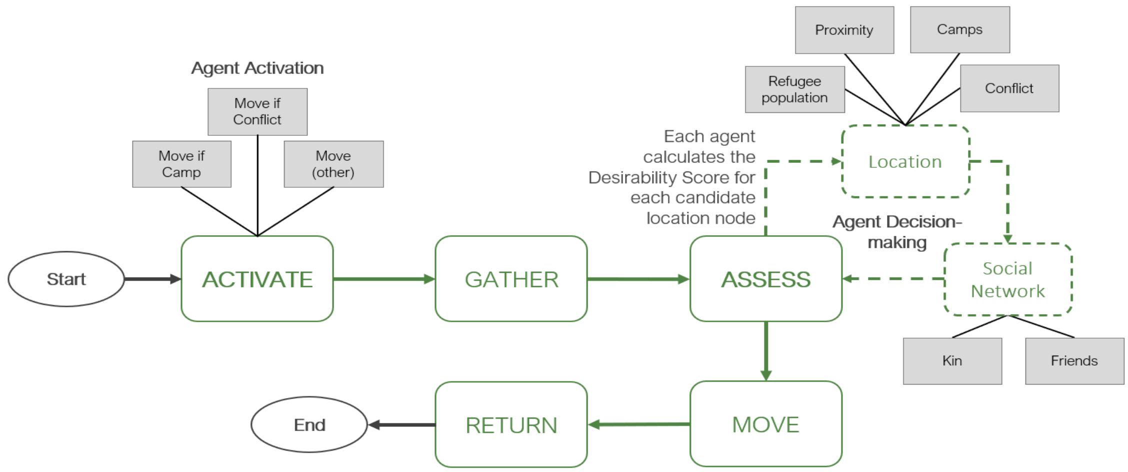


Figure 4 Agent decision-making logic overview and flow diagram depicting what occurs within the parallelized child processes and how agents reach move decisions. Both equations are noted within the image. Gray boxes represent.

In the subsequent parent process Aggregate, the resulting dictionaries from all child processes are merged back into the master simulation environment and the environment is thus updated, each location node now reflecting new refugee counts in accordance with the agent movement that has been executed during the parallel processes. The final parent process, Create, only applies if using more detailed social networking features not yet tested in this research. More information about these features is provided in Chapter 6, Conclusion and Future Work.

The simulation generates output in shapefile format. Assuming the write_step_shapefile parameter is set to True, a shapefile will be created for every simulation step with the current agent locations throughout the simulation expressed in the field REFPOP. The simulation also generates a validation shapefile which has fields additional to REFPOP, namely: valPop (the validation value representing real-world number of refugees in that administrative area at simulation end) and simEnd_nor (the Mean Absolute Error (MAE) calculated for every administrative area). For the error reporting presented in Chapter 4, additional statistics were calculated beyond those available in the validation shapefile.

Design Concepts

Basic Principles

The design concepts that influence the model are the ideas of voluntary vs. forced migration, push/pull models of migration (Lee, 1966; Dorigo & Tobler, 2010), herd behavior (Epstein, 2008; Tranos et al., 2012; Adhikari, 2013; Moore and Shellman, 2006; Moore and Shellman, 2007, Docquier et al., 2014), the theory of competing destinations (Fotheringham, 1981; Fotheringham, 1983; Chun & Griffith, 2011), the theory of intervening opportunities (Stouffer, 1940; Stouffer, 1960), utility maximization decision theory (Arentze, Kowald, & Axhausen, 2013; DeAngelis & Diaz, 2019), decision theory formulated from direct observation (Klabunde & Willekens, 2016), and SNT (Borgatti & Ofem, 2010). This model exclusively addresses forced migration resulting from global conflict events affecting the physical security and wellbeing of displaced persons. Most traditional migration models (e.g., spatial interaction models) are focused on voluntary migration and make extensive use of gravity models, where stocks and flows of migrants are calculated based on the geographic distance between two locations, the population sizes of each location, and any intervening or influential variables that might otherwise affect movement (Lewer & Van Den Berg, 2008). These models neither model migration flows at the individual level taking into account individual decision-making processes, nor do they address explicitly spatial patterns of flight beyond geographic origin and destination locations, typically at the country level (Edwards, 2008). For this reason, ABMs are an ideal tool for modeling and predicting forced migration patterns as they provide a mechanism to observe emergent spatial patterns from individual-level action. The basic principles applicable to this modeling method of forced migration include:

Push/pull factor theory: the idea that migrant mobility is influenced by factors that both repel a migrant from a location (push) and attract a migrant to a location (pull) (Lee, 1966; Ravenstein, 1885; Dorigo & Tobler, 2010). This model incorporates both push factors (e.g., conflict events) and pull factors (e.g., official refugee camps, existing refugee population).

Herd behavior and Social Network Theory: the idea that migrants will move to locations where they know others of the same sociocultural background have gone or are currently located (Epstein, 2008, Epstein & Gang, 2006; Tranos et al., 2015; Havinga & Bocker, 1999; Robinson & Segrott, 2002; Adhikari, 2013; Moore & Shellman, 2006; Moore & Shellman, 2007; Docquier et al., 2014; Rehm, 2012). Herd behavior appears in this model in dynamic pull factors. The factors are the presence of existing refugee populations, official refugee camps established by aid organizations, and members of an agent's social network in a candidate migration location. The way the social network is specified within the model lends itself to the inclusion of Grannovetter's (1977) seminal theory of the strength of weak ties within the broader domain of SNT. Given that agents form social ties randomly to both those agents who are entering the simulation at the same time as a target agent and to agents who are pre-existing in the simulation environment, this represents refugee social connections derived from both familial links as well as newer connections made, for example, online, through social media, or through networking groups (Borkert, Fisher, & Yafi, 2018; Dekker et al., 2018; Richey, 2014b).

Theory of competing destinations: the idea that the relationship between a migrant's current location and candidate future location must be contextualized within the possibility of migrating to all other candidate locations (Fotheringham, 1981; Fotheringham, 1983; Chun & Griffith, 2011). The theory of competing destinations appears in this model insofar as migrant agents calculate candidate location desirability scores for all candidate node locations in the migrant's view before making the decision to migrate to the most desirable location (or, as the case may be, remain in the current location).

Theory of intervening opportunities: the idea that the key driver of migrant decision-making is opportunities available at the end destination, not any factor related to the length or potential danger of the journey itself, or other opportunities that might present themselves along the way (Stouffer, 1940; Stouffer, 1960; Effers et al., 2008). In keeping with this theory, the simulation environment and agent decision-making are agnostic to both distance of transit and means of transportation. Additionally, through structured interviews, it was revealed that refugee migrants originating in Syria and moving to Turkey leveraged a variety of modes of transportation (to include car, on foot, boats, buses, and, least commonly, air travel) without favoring one over the other (Richey, 2014a; Richey, 2014b).

These are the logical theories and basic principles that affect simulation design, per Gilbert and Troitzsch's (2005) example. The decision theories that affect agent decision-making are discussed in the Objectives section of this ODD.

Emergence and Adaptation

Ideal social simulations will provide multiple paths to stochastic and probabilistic output, not singular paths to deterministic output (Cioffi-Revilla, 2005). In this vein, "emergence is an essential characteristic of social simulation," without which a simulation may not actually be considered *social* (Gilbert, 2002, p. 1). Inherently, emergent properties of social simulations are not reducible to individual action though it is the aggregate of individual behavior that produces the emergent phenomena. (Sawyer, 2001; Gilbert & Troitzsch, 2005; Miller & Page, 2009; Simon, 1996).

Emergence is largely derived from the stochastic or probabilistic elements of a simulation – in the case of this ABM, the social network and the dynamically updating location of agents throughout. The creation of the social network in this model is random on model initiation. Hinsch and Bijak (2019) found migration routes to be an emergent property of agent-to-agent communication. In this model, decision-making and movement patterns are heavily influenced by the strength of the migrant's social network, the relative sizes of those networks, and the respective geographic dispersion of those networks across the spatial extent, resulting in varying patterns of individual and collective agent movement. With this design, Hinsch and Bijak's (2019) findings will be considered replicated if the spatial patterns of migrant movement when using social networks differs greatly from the spatial patterns generated without social networks.

Objectives

Klabunde and Willekens (2016) contend that the agents or classes of agents in every simulation should adhere to some structured theory of decision-making, and further evaluate six types of decision theories for use within ABMs of forced migration. Most notably, the decision theory "should allow for the possibility that there is a gap between desires or intentions and actual behavior" (Klabunde & Willekens, 2016, p. 78). Gray, Hilton, and Bijak (2017) provide a very in-depth narrative around three elements that should play an important role in agent decision-making: time, uncertainty, and heterogeneity of decision logic. They argue that the more premeditated the migration event, the more extended the time horizon should be over which agents make decisions, the implications of which being that forced or reactive migrant decision-making should occur on the shortest time scale. The work of Dekker et al. (2018) further outlines the uncertainty migrants face during forced migration events, the lack of trust in both official and unofficial information sources, and the reliance on social media and the Internet to make decisions about where to go and how to arrive at the intended destination. Candidate decision theories for forced migrant agents include:

Utility or reward maximation: Agents conduct behavior that results in maximum reward as defined by the simulation objectives (Arentze, Kowald, & Axhausen, 2013). While utility maximization may appear logical for forced migrants, Reinhardt et al. (2019) have shown that forced migrants often make sub-optimal decisions due to social influence; in other words, decisions that would not always maximize utility or reward.

*Theory of Planned Behavior or Theory of Reasoned Action***:** Agents formulate attitudes towards certain behaviors largely based on the assessed probability of occurrence. From these attitudes, subjective norms, and perceived behavioral control, agents form intentions, and then enact behavior with some randomness (Ajzen, 1981; Ajzen, 1985). The theory of planned behavior, as its name implies, lends itself to strategic or reasoned action that occurs on a longer time scale than is appropriate for modeling forced migrants. An example of how the Theory of Planned Behavior has been applied in the context of voluntary migration is available in Klabunde, Willekens, and Leuchter (2017).

*Fast and Frugal decision heuristics***:** Agents maintain simple rules-based cognition that is easily influenced or overridden by social influence (Gigerenzer & Todd, 1999; Gigerenzer & Gaissmaier, 2011).

*Decision theory formulated from direct observation***:** In many cases, the most accurate form of decision theory is the decision theory observed in empirical data collected through field research or surveys, with stylized facts the generated to reproduce the scenarios observed in the field (Klabunde & Willikens, 2016). This approach is relevant to migrant decision-making given the number of studies now available that provide results obtained from field surveys and structured of refugees (Haug, 2008; McAuliffe, 2013; Robinson & Segrott, 2002; Borkert, Fisher, & Yafi, 2018; Dekker et. Al, 2018; Maitland & Xu, 2015; Richey, 2014b).

*Machine Learning***:** Agent decision-making logic will evolve over time in accordance with behavioral machine learning (ML) models such as Neural Networks (NN) or Genetic Algorithms (GA) (DeAngelis & Diaz, 2019; Rand, 2006).

Hinsch and Bijak (2019) demonstrate, using a theoretical ABM of the impact of social networking and information exchange on migration routes, that migrants frequently settle on migration routes that are not optimal under present conditions due to SNT, meaning that Rational Choice theories such as the Theory of Planned Behavior or Theory of Reasoned Action may not be the most appropriate decision theories for forced migrants. Additionally, the Theory of Planned Behavior or Belief-Desire-Intention (BDI) frameworks in ABMs are most appropriate when the complexities governing the agent decision-making process are well-known (Wolfe, Sierhuis, & Jarvis, 2008). As social networks have only recently been considered in ABMs of forced migration, these more complex decision processes for the creation of intelligent agents are likely not ideal for forced migrant agents. The decision theory leveraged by migrant agents in this ABM is a combination of utility maximization theory and decision theory formulated from direct observation, operationalized as the dynamic calculations of node desirability score and stochasticity of agent decision-making as a result. Agents making decisions exclusively based on a deterministic location score in the model would constitute utility maximization – move to the candidate move location that provides the highest reward. In this model, however, utility maximization can be overruled by the stochastic elements of the agent's social network causing agents to choose a less optimal candidate move location for social reasons. This decision theory is informed by direct observation

(Richey, 2014b) and further substantiated by the recent work of Hinsch and Bijak (2019), Dekker et al. (2018), Borkert, Fisher, & Yafi (2018), and Maitland and Xu (2015).

The decision theory to which an agent adheres facilitates the agent working towards the overarching simulation objective(s). In the case of this ABM, the objective could be considered movement itself as a result of push factors, all of which represent the driving factor of forced migration as threat to physical and personal security. Given the duration of the Syrian Civil War, however, migrants have had years to make more strategic calculations regarding movement patterns, and, in some cases, to move initially and move again later. For example, many migrants initially moved into Turkey to escape immediate threat, and then made longer-term and more calculated plans to migrate into Eastern and Western Europe (Baban et al., 2016; Yaylaci & Karakus, 2015). Given the popularity and perception of Europe as an ending destination, assuming migrants are unwilling to return home to Syria, a variable was explicitly included in the simulation representative of this overarching agent objective. The Location variable, currently set at model initiation to the geographic centroid of London, UK, is an exogenous factor that allows for the endogenous ranking of each candidate move location based on its proximity to a migrant's ending objective of reaching Europe.

Sensing

The implementation of agent sensing is simple. Agents are aware of all environmental variables associated with all neighboring nodes one degree of separation from the agent's current location. This awareness is provided in terms of the node location score and is in keeping with the theory of bounded rationality in agent awareness

(Edmonds, 1999). Agents also know the geographic locations of all their kin and friend agents within the simulation environment. This awareness, supported through direct observation and conversations with Syrian refugees in Turkey, is to replicate the effects of social media and virtual communication platforms such as Facebook, WhatsApp, and Instagram, all of which play host to informal migrant and refugee community pages that refugees use to assess candidate migration locations at a distance (Dekker et al., 2018; Borkert, Fisher, & Yafi, 2018; Brunwasser, 2015; Maitland & Xu, 2015; Richey, 2014b). Further, Simon (2019) has shown in an implementation of an ABM that a refugee operates in an incomplete information environment, represented computationally through bounded rationality, and resorts to a path dependence heavily influenced by social ties. This emergent path dependence is aligned with agent decision logic which satisfies rather than optimizes (Edmonds, 1999).

Interaction

Agent interaction in this ABM is mediated, meaning that agents interact with one another indirectly via the simulation environment. For example, an agent present in another agent's social network moves and therefore updates the simulation environment with its most current location. The other agent then uses this location to make its decision in moving.

Stochasticity

Stochasticity in this model is present in the dynamic social network variables and endogenous calculations at each time step. Each agent initializes with a random number of friends and a random number of kin. Each agent then moves probabilistically throughout the simulation based on its current environment and the locations of other agents, taking into account the locations of its social ties while moving. This dynamic and evolving social network among agents, the probability with which agents make the decision to move, and the random number of agents seeded at border crossings at each time step comprise the stochastic elements of the simulation, ensuring that the simulation predictions are non-deterministic and yield different results with every model run, even when calibrated to a specific real-world time period.

Details

Initialization

The model is initialized with the base variable values represented in Table 6.

Table 6 Model initialization variables and respective values.

Variable Name	Initialization Value	Description
data_dir	/data	Directory where data are stored
output_dir	/final_model_run	Name of run
draw_geo_graph	False	Print base map network
print_node_weights	False	Print node weights at each step
write_step_shapefiles	True	Write out shapefiles at every time step
num_steps	60	Number of simulation steps
num_batches	16	Number of batches to chunk refugees (typically same as number of CPU cores on machine)
num_processes	16	Number of parallel processes (typically same as number of CPU cores on machine)
num_friends	(1,5)	Integer or range; number of friendship ties to create for each refugee
num_kin	(1,5)	Integer or range; number of kinship ties to create for each refugee
camp_move_probability	0.7	Probability of movement if located in a node with an official refugee camp
conflict_move_probability	1	Probability of movement if located in a node with conflict
other_move_probability	1	Probability of movement if located in a node without camp or conflict
anchor_location	(51.5974, -0.1278)	Lat/Long of general direction of movement beyond spatial extent of simulation (can be NULL)
population_weight	0.3	Weight of pre-existing refugee population variable
location_weight	0.75	Weight of anchor location
camp_weight	0.5	Weight of presence of camp
conflict_weight	0.5	Weight of presence of conflict
kin_weight	0.1	Weight of kinship ties
friend_weight	0.1	Weight of friendship ties
seed_refs_per_node	(10,75)	Integer or range; number of refugees to seed at each border crossing location at each step
seed_nodes	Kilis, Yayladagi, Hatay, Akcakale	List of border crossing node IDs for seeding
new_friends_lower	0	Lower bound for the creation of new friendship ties when ties are created endogenous to the model (not used in base model)
new_friends_upper	0	Upper bound for the creation of new friendship ties when ties are created endogenous to the model (not used in base model)

The first part of model initialization is the creation of the simulation environment. Environment creation is part of the main python script but is also separated out into a second python script available for creating a simulation environment for any AOI: pre_process_basic.py. The simulation environment script takes as input vector data representing the administrative divisions at the level of granularity required for the simulation environment stored in the 'data' folder. The input shapefile will require at least four attributes:

1. Name or ID field containing the name or ID (string or integer) attributed to each individual location node.
2. Area field (a geometry field typical of all shapefiles)
3. Perimeter field (a geometry field typical of all shapefiles)
4. REFPOP field containing the pre-existing refugee population for each administrative area at time of real-world simulation initiation

Upon opening the preprocessing script, the script can be pointed to the input shapefile on the local or virtual machine. New shapefiles will be written out to the working directory and can be visualized using QGIS or other geospatial visualization software. The simulation environment will display after the script runs as specified.

For clarity, a simulation environment for CAR is provided below as an example. There are three primary administrative levels for CAR: administrative level 1 is the spatial extent of the entire country (n=1), administrative level 2 contains all prefectures which are similar in spatial extent to Turkish provinces (n=16 vs. n=81), and administrative level 3 contains subprefectures which are similar in spatial extent to

Turkish districts (n=48 vs n=929). The simulation environment for CAR, then, can be created with 16 or 48 location nodes respectively depending on the administrative level that the input shapefile represents. The two contrasting simulation environments are presented in Figure 5 and were created using the pre_process_basic.py script. These computational tools are provided in response to RQ2.

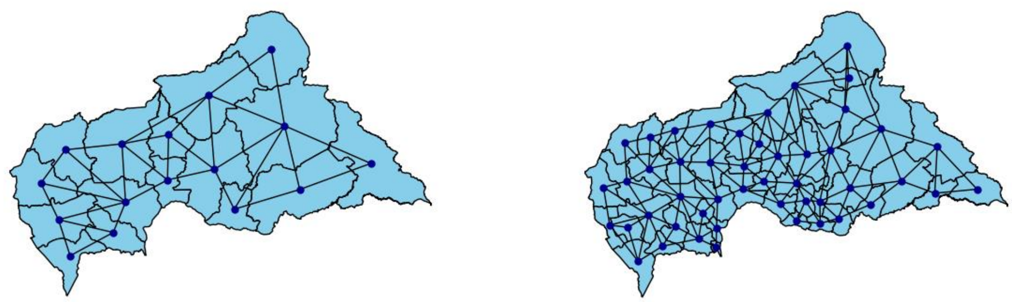


Figure 5 Two simulation environments for Central African Republic (CAR) with different granularity: 16 location nodes (left) and 48 location nodes (right).

The model is calibrated to the real-world time period the simulation is designed to represent – this may be backdated to the past, or present day running into a future state for prediction. Calibration was achieved through sensitivity testing using the static simulation environment variables as reported in Chapter 5, Case Study, Verification and Validation. Briefly, a feature scaling technique was applied to the output of a controlled model run with base parameters. The same technique was applied to the output of many model variations both isolating specific model parameters to identify thresholds and combining model parameters in distinct ways to determine the ideal calibration. The normalized output of each model variation was compared both to the control model and to real-world validation data with accuracy quantified as the normalized Mean Absolute Error (NMAE) of a target model condition to the control. The initial simulation environment is augmented after initialization using input data for existing refugee camps, and security or conflict events.

Finally, agents are initialized in the simulation from the shapefiles which should contain the pre-existing refugee population in the AOI. If there are no pre-existing refugee populations in the simulation, this value does not need to be imported and the simulation will begin seeding with zero refugees in the AOI at the identified seed locations with the identified range or fixed number of agents. With this flexibility, the model can initiate with zero agents or millions. The social networks are formed randomly in the model with each agent initializing with a random or fixed number of ties, as specified in the model parameters, from pre-existing or recently seeded agents.

Input Data

The simulation environment is created organically using input shapefiles and open-source datasets. The open-source data sets used in this model are:

1. The UNHCR's current locations of refugee camps and open border crossings during an ongoing security crisis or event, frequently identified on the Humanitarian Data Exchange (HDX) portal or OCHA's Relief Web (UNOCHA, 2020; HDX, 2020)[1]

---

[1] https://data.humdata.org/dataset

2. The Armed Conflict Location & Event Data Project (ACLED) dataset for the appropriate time period filtered down to include only violent protest, terrorist, or other security-related events in the simulation spatial extent (Raleigh et al., 2010)[2]
3. The UNHCR's (or other source's) existing refugee population by some administrative level above administrative level 0 (UNHCR, 2020)
4. Open-source shapefiles representing the spatial extent of the simulation environment at the desired level of granularity for the simulation[3]

The model must be temporally aligned to the target time period using these datasets because, for example, border crossings open and close and refugee camps are established and taken down by a variety of aid organizations. As such, ACLED, UNHCR, and other data must be pulled and pre-processed within the simulation for the month the researcher wishes to begin the simulation, whether this is present day or backdated for validation purposes.

Submodels

There are no submodels associated with the model documented above.

[2] https://acleddata.com/data-export-tool/
[3] https://www.naturalearthdata.com/

4. IMPLEMENTATION AND COMPUTATION

The following is an adaptation of a paper accepted for publication and presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC) from 30 November 2020 – 4 December 2020. The work in this Chapter addresses RQ2. Section 4.1 addresses computational challenges and Section 4.2 covers code implementation details of this model. Section 4.3 details the parallelization techniques in the implementation of this model. Section 4.4 presents series of experiments that test the effects of parallelization on model runtime and Section 4.5 concludes with a discussion of the parallelization results.

4.1 Computational Challenges

As argued in Chapters 1 and 2, ABMs are the ideal type of model for modeling the patterns and flows of forced migrants. The development of such models has obvious practical implications for humanitarian crisis management (Frydenlund et al., 2018). Despite this powerful research potential, the use of ABM for simulation at scale – millions of agents – and inclusion of robust migrant social networks is still in its infancy. The ABMs that do exist for modeling forced migration and other sociocultural phenomena are increasing in size and scale, presenting an imminent need to model millions of individual agents over many simulation timesteps. Even with simple model logic and a minimal number of model parameters, the computational overhead quickly becomes unwieldy both in memory and processing power required. This computational overhead scales exponentially when social networks are included in agent decision logic.

Techniques have been proposed in the past that seek to aggregate groups of agents into meta-agents, known as agent compression (Wendel & Dibble, 2007) or the creation of super-agents (Parry & Bithel, 2012). Parry and Evans (2008) show, however, that these super-agents are sensitive to spatial tasks. Watts (2016) stipulates the need for modelers to test scale dependency in models through sensitivity testing to determine whether the model developed in this dissertation necessitates millions of agents. The sociocultural modeling and simulation community is also turning to the domain of high-performance computing (HPC) for parallelization methods such as multiprocessing and multithreading, with ABM packages such as MASON which facilitate social simulation at scale in Java (Luke et al., 2005). While the parallelization techniques and parallel computing experimentation described in this chapter and in Chapter 2 do not contribute to the broader parallel computing capabilities for running ABMs at scale (e.g., MASON), the parallel computing methods and results do offer insight into optimal computational architecture and instantiation for the forced migration model presented in this dissertation.

To further demonstrate that scalability is required for this simulation, Watts' (2016) advice was taken and a set of experiments of scale were each run for 60 time steps with 3,600, 36,000, 360,000, and 3.6M agents to determine if similar model results could be obtained simply through using a scaled down number of agents, alleviating the computational overhead required to process millions of agents simultaneously. The results of these experiments are provided in Table 7.

Table 7 Experiments of scale using different numbers of agents. Results demonstrate that model error is very high when fewer agents are used.

	3k agents	36k agents	360k agents	3.6M agents
nMAE	0.70	0.88	0.89	0.07

The results suggest that in order to replicate real-world results, the model necessitates realistic numbers of agents, in this case, millions, not thousands. The ending spatial distributions of the first three scale conditions were also drastically different from the two spatial distributions observed in the full model results available in Chapter 4. In both cases, agents clustered in random districts in the middle of the country rather than near the Syrian border, in urban centers, or in the northwest area near Istanbul. These results further suggest that refugee movement is an emergent social phenomenon that is not replicable at smaller scales, commensurate with the observations of Hinsch and Bijak (2019). The fact that the model necessitates millions of agents to replicate real-world patterns requires that scalable and parallel computational architecture be in place to support millions or tens of millions of agents in a single simulation.

4.2 Code Implementation Details

The model is written in Python 3 and makes use of the following packages: json, os, sys, csv, copy, random, math, unidecode, numpy, pandas, geopandas, networkx, matplotlib, multiprocessing, and GDAL/Fiona. The model has two object classes: Ref and Sim. The Ref class represents refugee agents and contains two methods for creating social networks. The Sim class contains four methods: run, which runs the simulation; step, which executes agent decision-making logic; process_refs, which tracks agent movement; and find_new_node, which calculates the desirability scores. The main simulation method also calls several miscellaneous methods that are responsible for preprocessing the data and creating the simulation environment, validating the simulation output, and running various computational experiments.

The main model script is run_abm_shared_mem.py and is available on the public GitHub repo: https://github.com/mrichey17/mig. While data preprocessing is included in the main simulation script, it is also available in isolation: pre_proces_basic.py.

4.3 Parallelization Techniques

The model is parallelized by batching refugees at every time step using the parallelization logic described in Section 3.2 ODD, Processing, Overview, and Scheduling. Notionally the number of agent batches and parallel processes should correspond to the number of CPU cores on the provisioned virtual or physical simulation machine but, as described previously, the number of agent batches can exceed the number of parallel processes should a researcher be interested in optimizing model run time further. An optimal agent batch size and batch number can be found through hyperparameter testing. As an alternative to batching, drip feeding processes with agents assures that no one process stalls processing a large batch of refugees while other processes sit idle. The downside of this is the large overhead in communicating new data to the processes, so agent batching is the preferred method here.

Once the number of parallel processes has been established, each child process receives a tuple containing start and stop indices with the agents for which the process is responsible. Each child process then works independently on its batch, pulling the agents into its distributed memory environment and aggregating the results back up to the parent process after the parallel computation is complete. By leveraging multiple processes, a 3.5X speed increase across experimental conditions is achieved through multiple computational experiments reported in subsequent sections.

4.4 Parallelization Experiments

The experiments in this section establish a baseline and identify thresholds for various parameters of the parallelization processes in the simulation. These include the effect the size and complexity of the simulation environment, number of refugee agents, and average size of an agent's social network have on simulation step time. The results of all experiments are averaged over 10 runs using 1,000 location nodes with an average of 5 location neighbors and 100,000 agents with static social networks containing 1 kin and 1 friend unless otherwise indicated. All tests are run on a VM with 104GB of memory, 16 virtual CPU cores, and use 4 parallel processes in Google Cloud Platform (GCP). For reference, an average full model run requires 30-90 time steps.

Table 8 Model step run times (in seconds) based on several structural factors: size of location network with agents that have both static and probabilistic social networks, location network with 10k nodes and varying numbers of adjacent location nodes, and number of agents in the simulation.

# Location Nodes (static social network)	Time per step	# Location Nodes (probabilistic social network)	Time per step	Avg. # Location Node Neighbors (10k nodes)	Time per step	# Agents	Time per step
10	3.68	10	5.52	1	2.47	100	0.13
50	3.75	50	5.62	5	2.55	1,000	0.13
100	3.87	100	5.76	50	3.41	10,000	0.42
500	4.46	500	6.24	100	4.51	100,000	3.42
1,000	5.00	1,000	6.91	1,000	26.41	1,000,000	31.86
10,000	17.43	10,000	19.25	5,000	144.66	10,000,000	330.96
100,000	243.67	100,000	243.79	10000	287.45	20,000,000	nd

Several experiments test how various structural elements of the model affect model run time. Table 8 presents the model step run time in seconds of the various conditions. The following structural elements were tested: effect number of simulation environment location nodes has on step time (refugees have a static social network with 2 ties), effect number of simulation environment location nodes has on step time (refugees have a probabilistic social network with between 0 and 6 ties), effect average simulation environment location graph density has on step time, and effect number of total model refugees has on step time. The results indicate two primary observations. First, model step time increases linearly with the number of agents in the model, averaging 30 seconds per million refugees while using 4 parallel processes. By contrast, serial processing with the same parameters averages over 10 minutes per million refugees using only a small, static social network for each agent. Second, it is number of agents in the network and number of social ties that most impact model step time, not size of the simulation environment or simulation environment location graph density.

Additional experiments quantify step time speedup across varying conditions with different numbers of parallel processes, as presented in Table 9.

Table 9 Scalability results from representative model runs.

# Agents	# Social Ties	# Processes	Time to Completion	Speedup
1M	2	1	34.13	1.00
1M	2	2	28.42	1.20
1M	2	4	19.78	1.72
1M	2	8	**16.04**	**2.12**
1M	**2**	**12**	16.40	2.08
1M	2	16	16.62	2.05
5M	2	1	181.50	1.00
5M	2	2	149.73	1.21
5M	2	4	104.53	1.73
5M	2	8	84.25	2.15
5M	**2**	**12**	**83.60**	**2.17**
5M	2	16	86.02	2.10
10M	2	1	347.60	1.00
10M	2	2	287.02	1.21
10M	2	4	189.76	1.83
10M	2	8	138.70	2.50
10M	**2**	**12**	**124.40**	**2.79**
10M	2	16	152.91	2.27
25M	2	1	918.89	1.00
25M	2	2	726.88	1.26
25M	**2**	**4**	**461.06**	**1.99**
25M	2	8	nd	nd
25M	2	12	nd	nd
25M	2	16	nd	nd
1M	50	1	249.61	1.00
1M	50	2	172.68	1.44
1M	50	4	106.55	2.34
1M	50	8	76.26	3.27
1M	50	12	72.26	3.45
1M	**50**	**16**	**70.46**	**3.54**

Parallel processing tests were run using 1 million, 5 million, 10 million, and 25 million agents comparatively, with a simulation environment size of approximately 1,000 nodes and a static social network size of 2 social ties per agent. With these conditions, we observe up to a 2.8x speedup in total simulation step time using 12 parallel processes and a 2x speedup using only 4 processes. In the final parallel test condition, we use 10 million refugee agents with a social network size of 50 ties per agent, which, using serial processing, would slow the simulation to unusable levels. In this condition, we observe a 3.5x speedup in total simulation step time. In even the most computationally intensive model conditions (i.e., millions of refugees with robust social networks of 50 other agents), speedup is significant and increases steadily with the more parallel processes that are used.

4.5 Discussion of Parallelization Experiments

Parallelization experiment results indicate a speed performance increase across all tested model conditions while multiprocessing with a batch size of between 2 and 16 parallel processes. Up to 6 parallel processes are typically available on a consumer-grade computing machine where up to 16+ parallel processes are possible when provisioned on VMs in cloud environments. This holds true for models with large volumes of refugee agents, models where refugee agents have larger social networks, and models with larger simulation environments than that currently represented in the full model condition. Speed, however, does not increase linearly for each process as the overhead time spent on inter-process communication (IPC) at some point counteracts the speed increase obtained through parallel processing of agents or agent batches. This phenomenon is observed in most model conditions in Table 10 around 12 parallel processes, where using 16 parallel processes does not produce additional speedup. In no model condition did using greater than 16 parallel processes generate any model speedup suggesting a parallelization threshold for ABMs of varying levels of decision logic complexity with under 25M agents.

IPC time includes the time to copy the simulation environment to each child process and to write results back to the parent process. It is a relatively straightforward to parallelize an ABM with a few thousand agents, but as the number of agents, location nodes, and average graph density (i.e., number of edges) increases, the memory requirements to store the model parameters becomes very large (10s – 50s of GBs). For each child process across which computation is divided, it requires an individual copy of the simulation environment or, at the very least, access to variables that determine the stochastic portion of the node desirability score (i.e., the social variables). Eventually, the latency introduced by this copy overhead outweighs the speedup from parallelization and the constraint becomes total machine memory. This result can be observed in the 25M agent condition in Table 10 where maximum speedup achieved is 2X before memory errors occur.

Groen (2018) reports speed increases ranging from 2.5x to 3.4x using up to four parallel processes for the FLEE model of forced migration. These results replicate Groen's results with a 2x to 3.5x speed increase across several experimental conditions using between 4 and 16 parallel processes. While replicating the state-of-the-art ABM speedup results, this model also introduces static and stochastic elements not found in the

FLEE model, most notably a computationally intensive social network component. As shown in the initial experiments in Table 9, this is the most computationally intensive component of the model and therefore places the highest demands on parallel computing processes. The model also makes use of a simulation environment roughly 20 times the size of the FLEE model.

5. CASE STUDY

Chapter 4 introduces the case study where the model described in Chapter 3 was applied, verified, and validated. The case study in question is Syrian migration into Turkey as a result of the ongoing Syrian civil war conflict event. The work presented in this chapter addresses RQs 1 and 3. In Section 5.1, a brief history and contextualization of the conflict are provided. In Section 5.2, results for two model runs are reported, and verification and validation of those model conditions are reviewed in detail in Section 5.3. Section 5.4 provides a direct, one-to-one comparison of this model with the state-of-the-art model and Section 5.5 presents two other supporting case studies. Section 5.6 concludes this chapter with a discussion of all the observed results.

5.1 Introduction to Syria-Turkey Case Study

The Syria crisis has been ongoing since 2011 when political events subsequent to the Arab Spring escalated to violence. Sunni opposition groups called for the removal of Bashar al-Assad and his Shiite government and the events turned violent in March of 2011. Since that time, upwards of 7 million people have been displaced within Syria's borders and approximately that many outside of Syria's borders, over half of which now reside in Turkey. Other countries of asylum include Lebanon, Jordan, and Iraq. From 2011 to the date of writing, Syrian refugees have been migrating across the Turkish border either for asylum in Turkey itself or to transit Turkey en route to Eastern and Western Europe. At any given time, there have been 10-20 official border crossings open along the Syria/Turkey border maintained by officials from either country, though refugees often cross at unofficial border locations. Some are apprehended by Turkish authorities and returned to Syria but some make it through and go on to seek asylum in Turkey undocumented. In January 2014, I traveled to Turkey with an unofficial Syrian guide fluent in Turkish, Kurdish, and Arabic. We traveled along, with his sister, from Istanbul down to just south of Kayseri identifying unofficial refugee populations and conducting structured interviews, asking questions pertaining to their origin, their transit routes, their destination, their decision-making, their means of transportation, and their intentions to return to Syria if at all (Richey, 2014a; Richey, 2014b). Of the approximately 150 refugees interviewed, the use of social media, smartphones, and social connections to keep in touch with familial ties as well as make new connections along the way was paramount. Along with serving as primary motivation for this research, this experience was informative in creating an ABM with appropriate parameters to address the Syria/Turkey case study.

5.2 Results

The two final representative runs of the model addressed Syrian refugees crossing the Turkish border and moving throughout Turkey beginning on 1 February 2019 and ending on 30 March 2019. The first model condition, the ‘full’ model, represents the model inclusive of its social network mechanics. The second model condition, the ‘base’ model, represents the model without these mechanics. The two models are run for comparative purposes in satisfaction of RQ1. Both models are validated using the UNHCR’s data from 31 March 2019 which details the refugee population, both within and outside of official refugee camps, at the district level or administrative level 2. The models are initiated using the parameter values indicated in Section 3.2, Initialization. The models are run within GCP Cloud AI Jupyter notebooks on virtual machines provisioned with a 16 virtual CPU cluster, with 200GB of memory, running 16 parallel processes. The average total cost of the final model runs is $37.00, and the average final model runtime is 12.5 hours, or 96+ hours without using parallelization. All error reporting in this dissertation is based on the mean of 10 model runs. The error for both simulation conditions is presented in Figure 6.

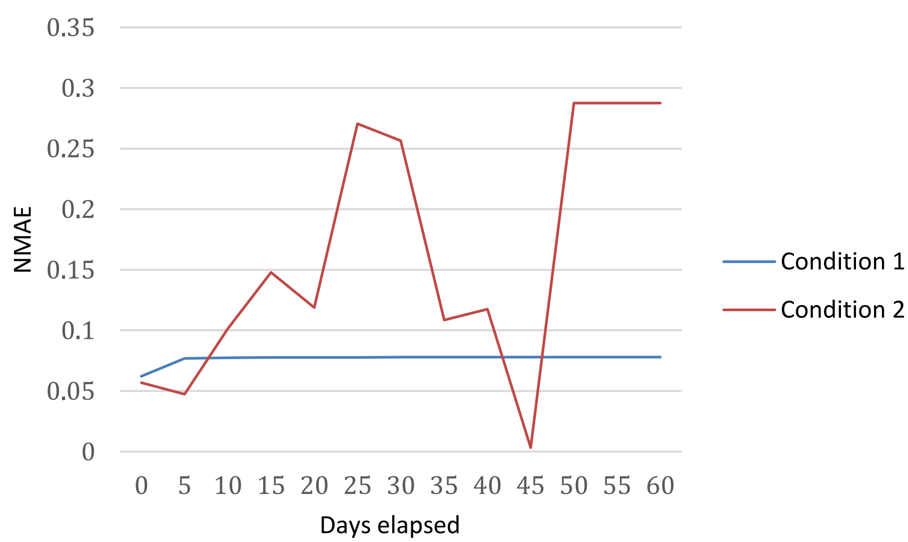


Figure 6 Error over time for both simulation conditions, representing full simulation runs with social networks (blue) and without social networks (red). Average error, measured as NMAE, is 0.07 for the condition with social networks and 0.11 for the condition without social networks indicating that the inclusion of social networks reduces model error.

Condition 1 – Including Social Networks

For the full model, inclusive of the social network mechanics, the number of refugees in-country at model initiation is 3,603,811 and the number of refugees after 60 days at the end of the simulation is 3,615,387, which is an addition of ~11,500 refugees in 60 days. The number of actual refugees in Turkey at simulation end, per the UNHCR, was 4,074,693. Model error for this condition is 0.07 as depicted in Figure 6, calculated as average NMAE over time per the description of error calculation below in Section 4.3. Using this statistic, it is clear that this model predicts forced migratory movement with an approximate error of 0.07 in satisfaction of RQ3, as depicted in Figure 6.

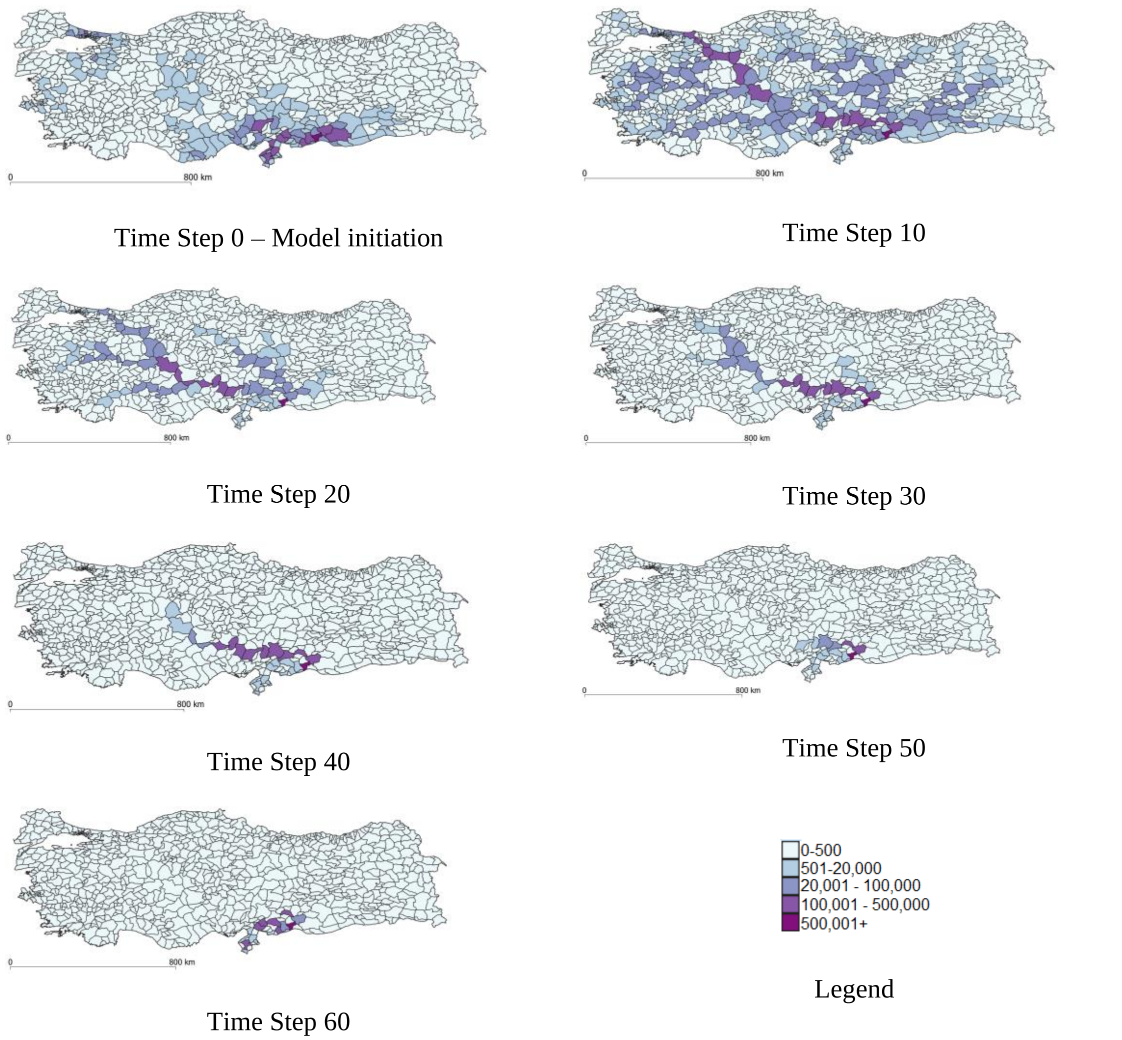


Figure 7 Spatial distribution of refugee agents every five time steps of the model condition with refugee social network mechanics. Refugees initially disperse correctly then retreat to the border crossing locations over time.

Initially, refugees are dispersed throughout the country and then begin to follow corridors of migrant flow which spatially represent actual patterns of refugee movement as depicted in Figure 7. As the simulation progresses, however, refugees begin to move away from the northwest quadrant of the simulation (which contains the largest urban center, Istanbul) back down towards the Syrian border. Sensitivity testing, explained further in Section 4.3, reveals that this pattern is likely due to the social network component in the model as this pattern is strongest when the social network component (represented through friendship and kinship ties) is active.

The initial specification of social influence in the model is naïve assuming random, bidirectional ties which generate reciprocal decision logic. Social ties – both friendship and kinship – are initiated at random for each refugee in the simulation as described in Chapter 3. This means that refugees already in the simulation in the northwest corner randomly generate ties to refugees who are near the Syrian border or further east in Turkey. It is likely acceptable that the ties themselves are bidirectional in that if Refugee 27 is friends with Refugee 30, Refugee 30 is also friends with Refugee 27. Reciprocity, however, should likely not manifest as reciprocal model logic, i.e., how refugees act on or use their social networks in deciding where to migrate. In practice, what occurs in the model is that refugees initiated in the northwest corner of the simulation environment actively seek to move closer to their kin and friends near the Syrian border. This results in a reverse migration of refugees towards the Syrian border. In reality, while refugees in Istanbul may have kinship and friendship ties closer to the border (in fact, it is known this is true), they are less likely to move towards these contacts, waiting instead for those contacts to migrate to their present location. That is the essence of the social network mechanic: that refugees in border regions likely have known (familial) or unknown (social media) contacts in other parts of the country that inform their decision-making and destination, but refugees that are already in camps, urban centers, or otherwise safe spaces do not necessarily move towards known or unknown social contacts even if these ties exist. This disconnect provides the most likely explanation for long term model inaccuracy and levies the requirement for two classes of agent refugees: one for those refugees who have recently crossed the border or seeded in the simulation environment, and one for those refugees who have been in the simulation environment for a while. Each class of refugee agent could then implement different logic that would emphasize or de-emphasize the social network in the decision to migrate accordingly. The implementation of multiple agent classes as well as variations on the theme of additional social network models are discussed further in Chapter 6.

Condition 2 – Excluding Social Networks

For the model, exclusive of the social network mechanics, the number of refugees in-country at model initiation is 3,603,811 and the number of refugees after 60 days at the end of the simulation is 3,606,422, with an addition of ~2,600 refugees during the course of the simulation. The number of actual refugees in Turkey at the time of simulation end, per the UNHCR, was 4,074,693. Model error for this condition is 0.11 as depicted in Figure 6, calculated as average NMAE over time per the description of error calculation below in Section 4.3.

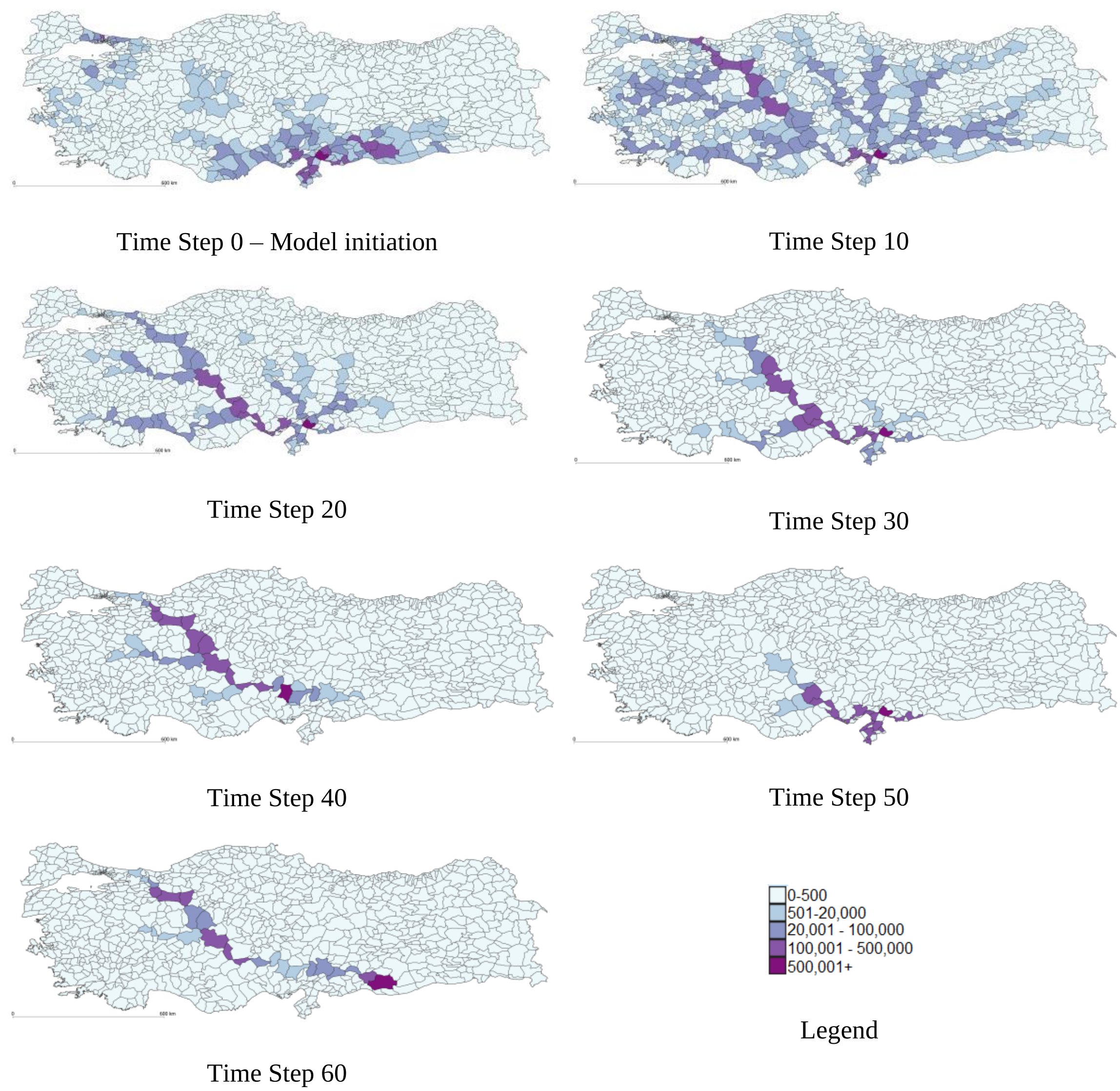


Figure 8 Spatial distribution of refugee agents every five time steps of the model condition with refugee social network mechanics. Refugees initially disperse correctly then retreat to the border crossing locations over time.

Model error is slightly higher in this condition where social networks are excluded and the spatial distribution of refugees at the end of 60 time steps is revealing. Where in Condition 1, it is evidence that the agents are collecting in certain areas due to social network mechanics, in this condition without social networks, higher path dependence is observed because agents are using structural features of the simulation environment to devise migration routes. The districts producing the majority of the error in Condition 1 are the central and northwesternmost districts near Istanbul and Izmir where refugee populations exist but the model does not replicate them. In the second condition, after removing the social network component, the agents move more freely throughout the simulation and disperse much further into the country when unconstrained by remaining close to social ties seeding at districts with border crossings. The error, then, comes from the variability in predicted and actual refugee populations, as depicted in Figure 6 because, without social ties, the refugees also do not naturally migrate to urban centers or refugee camps. In response to RQ1, it is observed that the inclusion of social networks serves to reduce model error but produces spatial patterns that indicate the implementation of social network modeling mechanics should be developed further, as discussed in Chapter 5.

5.3 Verification and Validation

Verification

As the first step of verification, a code walkthrough was performed with a qualified, master's-level computer scientist and software developer. The code was reviewed line by line to ensure continuity of the codebase. Revisions were also made accordingly to consolidate lengthy lines of code that could be represented in single lines using lambda functions and similar techniques. The amount of for and while loops were also reduced, and global variables separated and stored upfront. Code test conditions were also added to provide baselines and benchmarks for model runs assessing computational speedup and model accuracy.

Validation

To give the reader a clear understanding of the validation process, Table 7 provides an example of a simulated (predicted) refugee population for three hypothetical districts which are compared to the exemplar real-world refugee population in those same districts. To obtain model error, the simulated population is subtracted from the real-world population for each district, yielding the absolute value of the error. The absolute error is then normalized between 0 and 1 using the Min-Max Normalization feature scaling technique. NMAE is then calculated for each time step using the mean absolute error across all districts. This provides an NMAE score for each time step of the simulation. The average of NMAEs across the entire temporal window is considered the NMAE for the whole model. This can be done at the end of each simulation time step or at the end of the simulated run time. In the example provided in Table 10, the average model error is .35, though no averaging is required because this example assumes a single time step.

Table 10 Exemplar simulated and real-world data used to illustrate method for calculating simulation accuracy.

	District 1	District 2	District 3
Simulated refugee population	785,627	21,524	8,780
Real-world refugee population	880,951	19,723	1,857
Absolute Error	*95,324*	*1,801*	*6,923*
Absolute Error (normalized)	*1.0000*	*0.0000*	*0.0548*
Model NMAE	0.35		

Equally as important as the model's statistical error is the spatial distribution of that error since the model's main purpose is to predict the spatial distribution of refugees during a forced migration event. For this reason, it is useful for analysts and researchers to map the NMAE of each district to understand the spatial distribution of the deviations from real-world data as depicted in Figure 9.

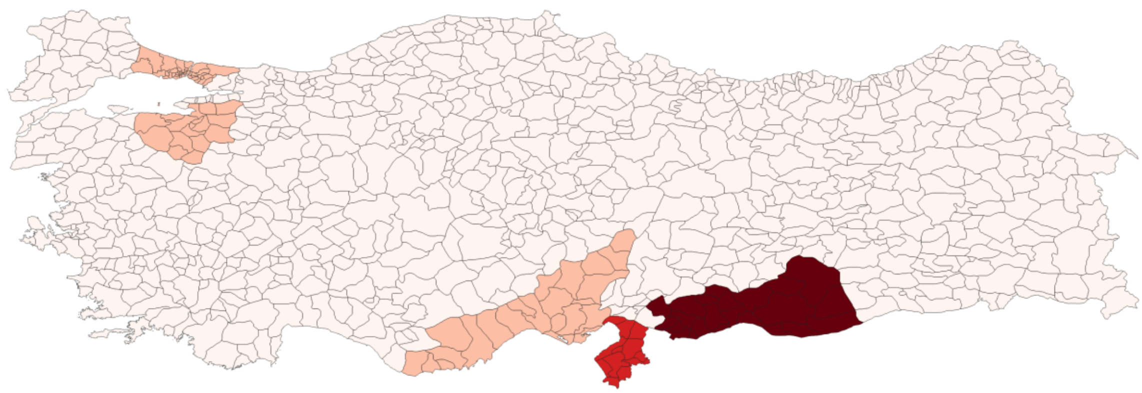


Figure 9 Exemplar Normalized Mean Absolute Error for each district within the full model condition after 60 days, measured between 0 and 1 and represented geospatially. A score of 1 in a district indicates significant deviation from actual refugee populations at time of simulation end, where a score of 0 indicates no deviation from real-world conditions. Districts with a score of 1 are colored dark red.

Additional validation and calibration of the model were achieved through robust sensitivity testing in keeping with Gilbert and Troitzsch's (2005) methodology, numerous model runs of varying lengths, and analysis of model output at each time step. 29 representative experiments, each run for 14 time steps, were selected to demonstrate the influence all model parameters have on model performance in addition to the thresholds for each parameter, displayed in Table 11.

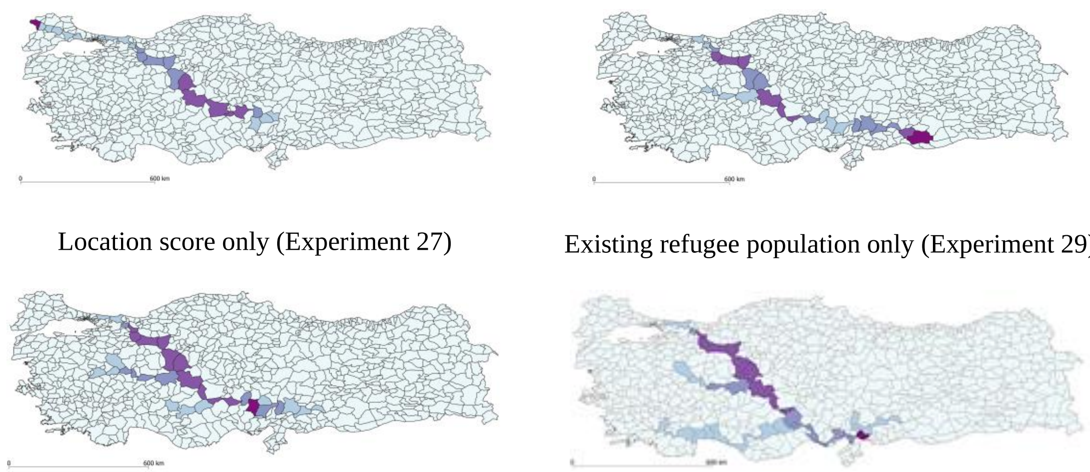
Table 11 Results from several experimental conditions used for model verification. Experiments were run for 14 time steps 10 times each with values averaged over these 10 representative runs.

Experiment Number	Experiment Name	Experiment Description	NMAE	% change from base accuracy
0	Base Parameters	Parameter set used to establish baseline against which to compare all experimental conditions	0.004220	NA
1	Friends & Kin 1	Generate between 0 and 5 friendship and kinship ties per refugee	0.005770	0.367286648
2	Friends & Kin 2	Generate between 5 and 10 friendship and kinship ties per refugee	0.003708	-0.357373859
3	Camp Move 1	Probability of moving if located at a refugee camp set to 70%	0.003715	0.001741822
4	Camp Move 2	Probability of moving if located at a refugee camp set to 100%	0.003671	-0.011792959
5	Conflict Move 1	Probability of moving if located in a conflict zone set to 30%	0.004220	0.149537324
6	Conflict Move 2	Probability of moving if located in a conflict zone set to 70%	0.004216	-0.000753598
7	Other Move 1	Probability of moving in the simulation set to 100%	0.003806	-0.097276539
8	Other Move 2	Probability of moving in the simulation set to 30%	0.006053	0.590212418
9	Seed Refs 1	25 refugees seed at each border crossing at each time step	0.004216	-0.303381824
10	Seed Refs 2	75 refugees seed at each border crossing at each time step	0.004213	-0.000805205
11	Seed Refs 3	100 refugees seed at each border crossing at each time step	0.004214	0.000219789
12	Seed Refs 4	500 refugees seed at each border crossing at each time step	0.004196	-0.004386259
13	Anchor Location Warsaw	Anchor location moved from London to Warsaw	0.004215	0.004535084
14	Anchor Location Moscow	Anchor location moved from London to Moscow	0.004213	-0.000308747
15	Population Weight 1	Importance of moving to locations with pre-existing refugee populations set to 0.75	0.003418	-0.188781189

Experiment Number	Experiment Name	Experiment Description	NMAE	% change from base accuracy
16	Population Weight 2	Importance of moving to locations with pre-existing refugee populations set to 1	0.003686	0.078440596
17	Location Weight 1	Importance of moving to locations closer to the anchor location set to 1	0.005047	0.369336618
18	Location Weight 2	Importance of moving to locations closer to the anchor location set to 0.75	0.005488	0.087348868
19	Camp Weight 1	Importance of locating a refugee camp set to 0.75	0.003915	-0.286656232
20	Camp Weight 2	Importance of locating a refugee camp set to 1	0.003912	-0.000864093
21	Conflict Weight 1	Importance of avoiding a conflict zone set to 0.75	0.004216	0.07775006
22	Conflict Weight 2	Importance of avoiding a conflict zone set to 1	0.004218	0.000478241
23	Kin Weight 1	Importance of kinship ties in a target location set to 0.75	0.013124	2.111715642
24	Kin Weight 2	Importance of kinship ties in a target location set to 1	0.013661	0.040898187
25	Friend Weight 1	Importance of friendship ties in a target location set to 0.75	0.013201	-0.033710609
26	Friend Weight 2	Importance of friendship ties in a target location set to 1	0.013703	0.038041608
27	Location Weight 3	Emphasize location score above all else (location score set to 1, all other parameters set to 0)	0.013190	-0.037420822
28	Social Network 1	Emphasize social network (friends and kin) above all else (friends and kin score set to 1, all other parameters set to 0)	0.003239	-0.754415182
29	Population Weight 3	Emphasize pre-existing refugee population above all else (population weight set to 1, all other parameters set to 0)	0.003686	0.137898756

There are several primary model parameters that affect agent movement throughout. They are the model's anchor location, the pre-existing refugee population, official refugee camps, conflict events, the social network (comprised of friends and kin), and the weights of each of these parameters respectively within internal model calculations. All experiments are run for 14 simulation steps (2 real-world weeks) and compared against Experiment 0 using a feature scaling technique. Min-Max Normalization was used for calculating base accuracy as the objective function. Z-Score Normalization was also considered due to the final distributions of refugees within the country which had numerous outliers in border districts, but Min-Max Normalization was ultimately chosen so direct comparisons and calculations could be made to calculate aggregated simulation accuracy. Experiment 0 is a base parameterized model specified as follows: num_friends = 1, num_kin = 1, camp_move_probability = 0.3, conflict_move_probability = 1, other_move_probability = 0.75, anchor_location = London, seed_refs_per_node = 50, and all weights = 0.25 (population_weight, location_weight, camp_weight, conflict_weight, friend_weight, and kin_weight).

There are several experiments that show improvement (decrease) to overall model error. Notably, when other parameters are active in a model, emphasizing parameters that relate to social networks or other refugee agents in the simulation tend to reduce error (Experiments 2, 9, 15, 19, and 28). The same is true for the pre-existing refugee population, though the effect is not as extreme. This can be interpreted to mean that the model does, indeed, necessitate a variety of competing factors and adequate balance of each to replicate real-world results. Also, that inclusion of the stochastic model elements tied to agent social networks decreases model error and improves performance. Increasing the importance of official refugee camps in the simulation improves model accuracy (Experiments 19 and 20) though model error does not improve when movement exclusively to pre-existing refugee camps is emphasized (Experiments 2 and 3) as is the case in the FLEE model. This further substantiates the use of an organic spatial network as the base layer of the simulation environments as opposed to a spatial network consisting of only official refugee camps or urban centers. Spatial distributions of refugees from the final three experiments appear in Figure 9.



Location score only (Experiment 27) — Existing refugee population only (Experiment 29)

Social network only (Experiment 28) — Mixed Parameters

Figure 10 Spatial distribution of refugees after 14 time steps (2 weeks) isolating certain parameters. The highest accuracy spatial patterns occur in the mixed parameter mode run, but over time, the social network component (lower left and lower right) draws refugees towards the Syrian border.

Finally, experiments 1-12 focused on sensitivity testing, or establishing appropriate thresholds for all model parameters. From these experiments, it is evident that including conflict events in the simulation has little effect on model accuracy (Experiments 5 and 6). This is likely because the majority of the conflict events were near the Syrian border, in areas away from which refugees are already moving, or in the eastern part of the country, where few refugees tend to collect. Conflict events may be more relevant in other, less-stable geographies such as CAR. Finally, it is evident in experiments 13 and 14 that the specific location of the anchor location, while important to agent objectives and decision-making logic (Experiments 17, 18, and 27), does not impact model accuracy. This suggests that it is important merely to have an anchor location that represents the general direction of refugee movement (in this case, Europe). The anchor location will be trickier to implement in areas such as CAR where refugees and IDPs alike tend to move in all directions simultaneously without clear directionality. In these cases, the anchor location can be removed entirely. Additional models should explore establishing anchor locations based on urban centers or lack of conflict (as conflict is also likely to be higher in these areas as well). It is evident in experiments 9-12 that it does not matter how many refugees are seeded at each time step. A separate set of experiments tested the scalability of model results the results of which are available in Section 4.1.

5.4 Comparison with Other Models

As discussed in previous chapters, the model that is most comparable model to the model presented in this work is the FLEE model, an open-source, parallelized model of forced migration written in the Python programming language[4] (Suleimenova, Bell, & Groen, 2017). The FLEE model initially reported an overall 75 percent accuracy in predicting refugee movement to camps across three geographies. Accuracy is not well defined in the FLEE literature, though model error is calculated as the mean relative

[4] https://github.com/djgroen/flee-release

difference between simulated refugees in camps and actual refugees in camps (UNHCR data) at each time step. The initial paper reported an average model error of 0.5 across all test conditions that falls to 0.1-0.3 over time across several hundred simulated steps. In practical terms, this means that the FLEE model is not useful for making predictions in the first 30-90 days of a simulated period, the time period that is most critical for analysts, operators, and aid workers. In comparison, this dissertation reports an average model error of 0.07 across multiple models for the first 60 days of a simulated period. To establish a baseline and means of comparison between the two models, a one-to-one comparison of the two models was performed in the Turkey AOI. The FLEE model was constructed, parameterized, and run according to the tutorial available on the FLEE website and personal instruction from the FLEE model creators, Diana Suleimenova and Derek Groen. Below, the results of that side-by-side comparison are reported.

Simulation Environments

While the model documented in this dissertation uses a simulation environment comprised of 929 network nodes, the FLEE model, when implemented in Turkey, uses a simulation environment comprised of only 38 nodes, of which only 21 are available to refugees are destination locations. The FLEE model uses pre-existing official refugee camps as candidate destination locations within its model, which substantially constrains the locations to which refugees can freely choose to move and, as such, model realism as discussed in Chapter 2. A rendering of the spatial extent of the FLEE model as applied in Turkey appears in Figure 11. Other locations in the FLEE model include conflict locations, or locations in which refugees are seeded in the simulation, and towns, transit hubs through which refugees can move to access camps if they are too geographically distant.

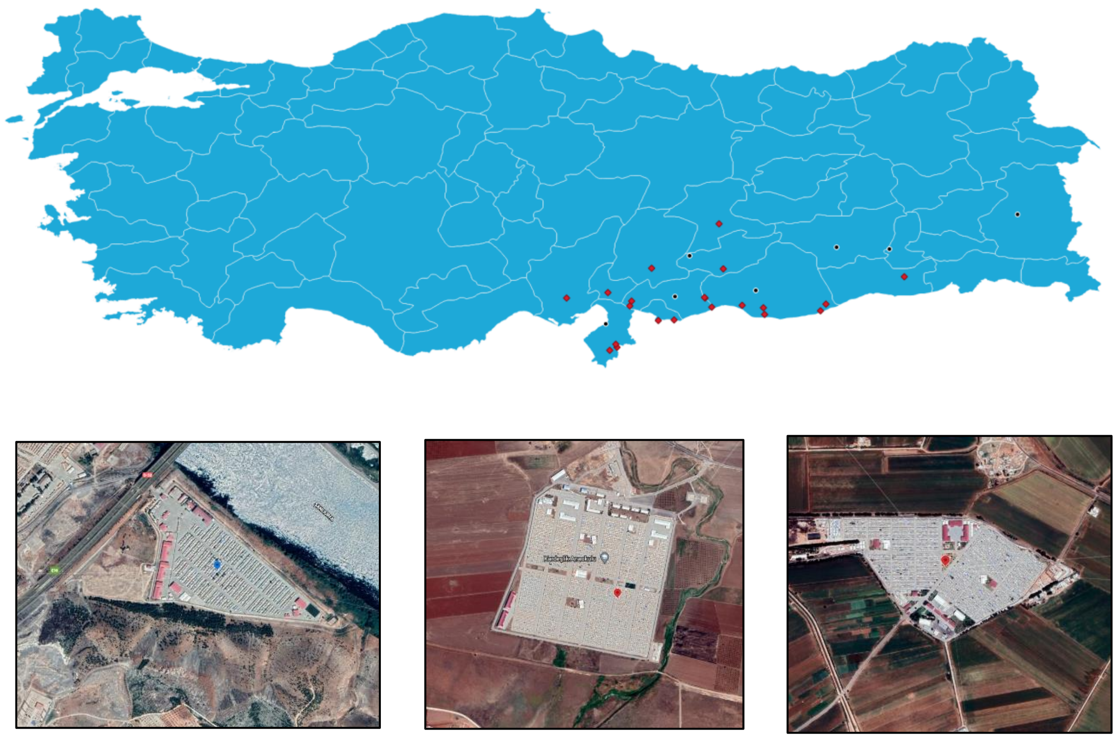


Figure 11 Input to the FLEE simulation environment (top) and official refugee camps as observed in satellite imagery (bottom). Red diamonds (top) represent pre-existing official UNHCR or governmental refugee camps. Black circles (top) represent transit hubs or cities.

All of the locations in the FLEE simulation environment and the network linkages between them must be created manually, by hand, by an analyst or researcher, which is both time-consuming and unrealistic for simulations of larger size. The simulation environment of the model presented in this dissertation is created from shapefile input and is generated automatically from the data pre-processing code, which creates both a larger and more organic simulation environment in a fraction of the time required to construct the simulation environment of the FLEE model as described in Chapter 5. A representation of the simulation environment for this dissertation's model is available in Figure 12. Both conflict zones and official refugee camps are still included in the simulation so no information is lost in the creation of the simulation environment using this method.

Figure 12 Network-based simulation environment of the model presented in this dissertation. Data pre-processing code automatically generates a simulation environment representative of any geography in under 20 seconds.

Given that approximately 98 percent of Syrian refugees reside in urban centers, it follows that using urban centers as location nodes is a viable option. Specifically, refugees in Turkey reside in three major urban centers: Istanbul, Izmir, and Ankara (International Crisis Group, 2018). Given the availability of data, however, and the prevalence of shapefiles at various administrative levels, the most flexible, realistic, and generalizable simulation environment employs the organic approach to creating a location network described above. If emphasis should then be placed on urban centers given the UNHCR statistic, administrative areas containing urban centers can be weighted more heavily in the simulation than areas without urban centers.

To create a one-to-one comparison of the FLEE model and the model presented herein, a model condition was created with the same spatial extent of the FLEE model. This model's simulation environment was created by importing the Turkey shapefile (administrative level 2, districts), and eliminating districts that did not contain an official refugee camp (the geolocations for which were already available from other data input of the simulation). The resulting simulation environment contains 48 nodes, 21 of which contain refugee camps, and is depicted in Figure 13. The 27 nodes in the simulation environment that do not contain official camps remain in the simulation to serve as transit nodes between the camps; these nodes serve the same purpose as the transit hubs in the FLEE model.

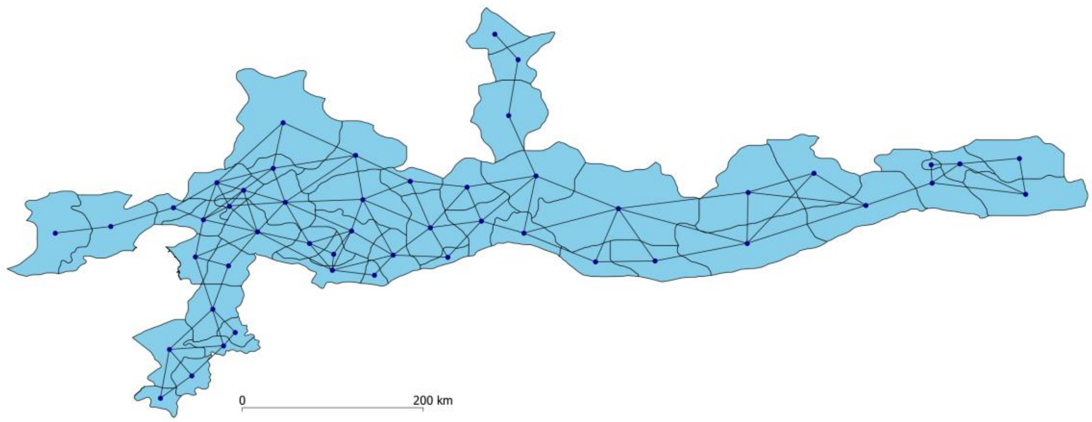


Figure 13 Simulation environment of the reduced simulation model. This spatial extent matches that of the FLEE model represented in Figure 7 but was created using the data pre-processing code relevant to this dissertation.

Results

The FLEE model's NMAE was 0.33 for the 60-day simulated time period compared with the reduced model's reported NMAE of 0.09 for the same time period, both of which appear in Figure 14.

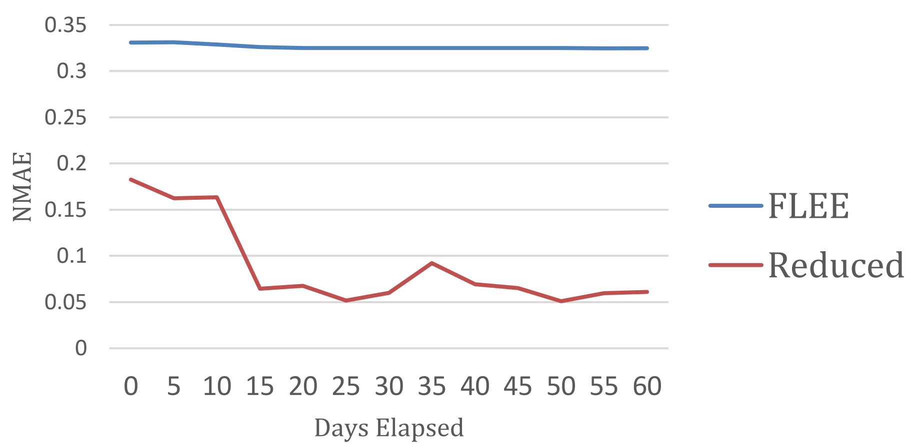


Figure 14 Error over time for the FLEE model and the reduced model with the same spatial extent as the FLEE model. Error, reported as NMAE, is 0.33 for the FLEE model and 0.09 for the model.

The spatial distribution of the refugees in the model is presented in Figure 15.

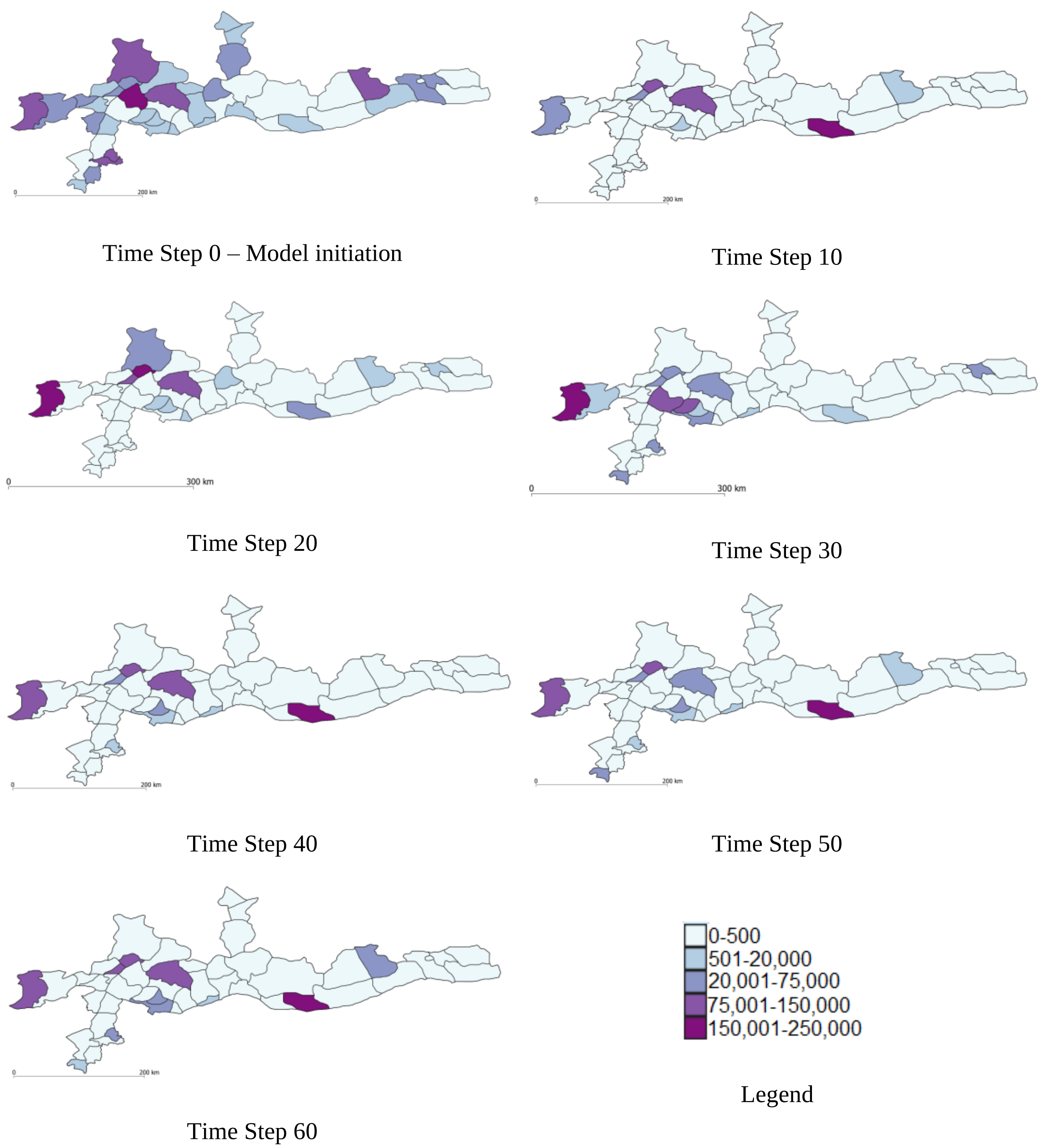


Figure 15 Spatial distribution of refugee agents every ten time steps of the model with reduced spatial extent.

5.5 Additional Case Studies

The Syrian Civil War affects six countries: Turkey, Lebanon, Jordan, Israel, Iraq, and Iran. Of these, the top three refugee hosting countries are Turkey, Jordan, and Lebanon (UNHCR, 2020d). To test the generalizability of this model across the broader Syrian conflict, the model was run in both Jordan and Lebanon with similar set up to the Turkey case study. Each country model comprised two test conditions: Condition 1 which includes social networks and Condition 2 which does not. The critical difference between Turkey and these other two countries during the study window (1 January 2019 and 31 March 2019) is that, during this time, Turkey was still being flooded with new refugees where the mass migration to both Jordan and Lebanon was in decline with many refugees already in those countries becoming returnees and returning to Syria (Sewell, 2020; Roggio, 2013; Eldawy, 2019). Additionally, there are no formal refugee camps in Lebanon per orders from the Lebanese government, so all refugee settlements are emergent (Cherri, Gonzalez, & Delgado, 2016; Blanchet, Fouad, & Pherali, 2016; Sewell, 2020; Hijazi, Lovatt, & Iraqi, n.d.). In both Jordan and Lebanon, there are also many Palestinian refugees and 12 Palestinian refugee camps in Jordan, all of which were already present in the country at the time of the Syrian crisis (Hijazi, Lovatt, & Iraqi, n.d.).

During the study window, Syria had five official border crossings into Lebanon (Jagarnathsingh, 2019) and Lebanon had 12 camps (Eldawy, 2019). Syria had one border crossing into Jordan (Roggio, 2013) and four known emergent refugee camps, which were used, in this case, in place of official camps (UNHCR, 2020f). The total number of

Syrian refugees in Lebanon on Day 1 of the study window was 948,849 (UNHCR, 2020g) and 702,970 in Jordan (UNHCR, 2020h). On both counts, these numbers are smaller than Turkey. The study areas for both countries were also much smaller and much less granular than the Turkey case study. To better assess the model's flexibility, the model was seeded with a fraction of the refugees in the case study country on Day 1 of the simulation instead of the actual number of refugees. The number of refugees coming across the border, then, simulates a flood of refugees that were then free to disperse throughout the country to see if the model was capable of predicting migration patterns not wholly reliant on pre-existing refugee locations. This was also necessary because, as previously mentioned, during the time of the study window, there were not many new refugees entering the country and refugees were even beginning to return to Syria – a situation for which this model does not fully account.

Lebanon

In Lebanon, Administrative Level 2 shapefiles were used which produced a location graph with 30 nodes, as depicted in Figure 16.

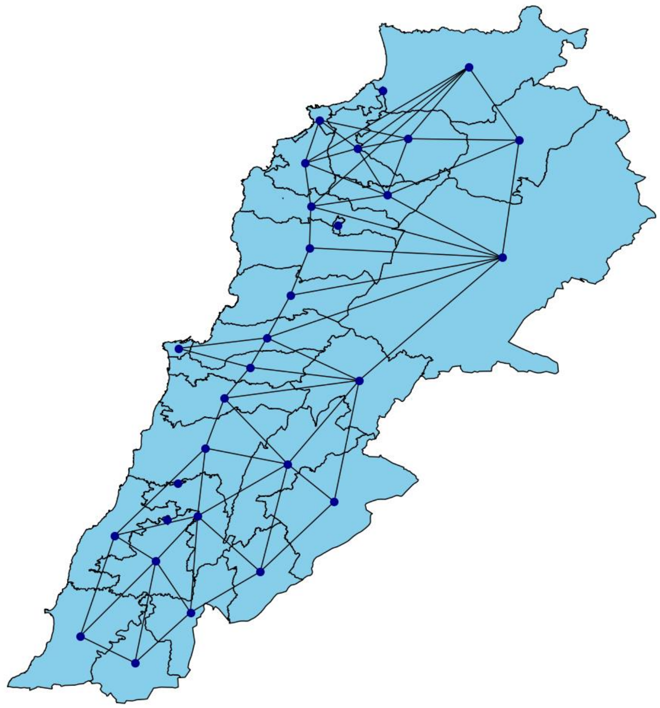


Figure 16 Location graph of Lebanon with 30 location nodes.

There were 12,092 refugee agents at initialization which grew to 1,294,855 refugees in Condition 1 (with social networks) and 1,301,061 in Condition 2 (without social networks). The actual number of refugees used to validate the simulation was 944,613 (UNHCR, 2020g) – this number is lower than the simulation numbers because approximately 4,200 refugees had returned to Syria during the study window. Model error (NMAE) is 0.15 for both conditions of the model, producing only a slight measurable difference between the condition with and the condition without social networks, per Figure 17.

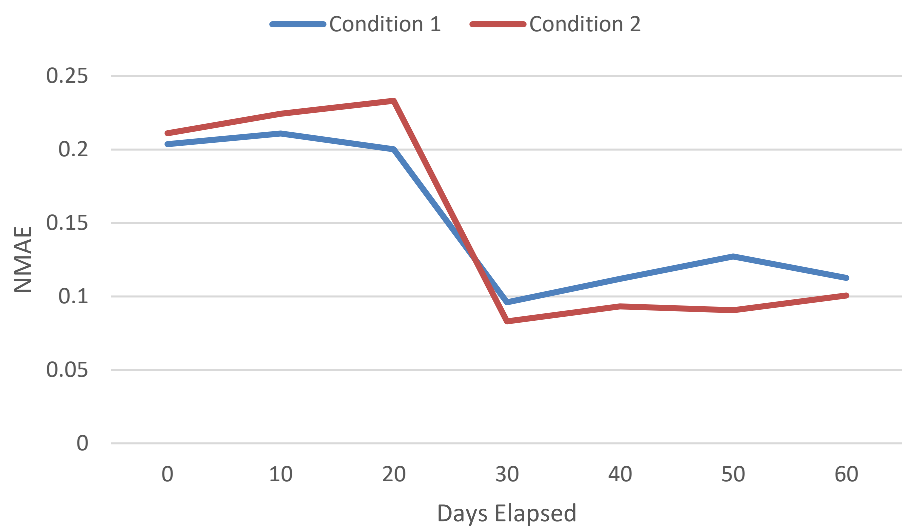


Figure 17 Error over time for the Lebanon case study. In both conditions, model error is 0.15.

The spatial distribution of refugee movement throughout the simulation in Condition 1 is available in Figure 18.

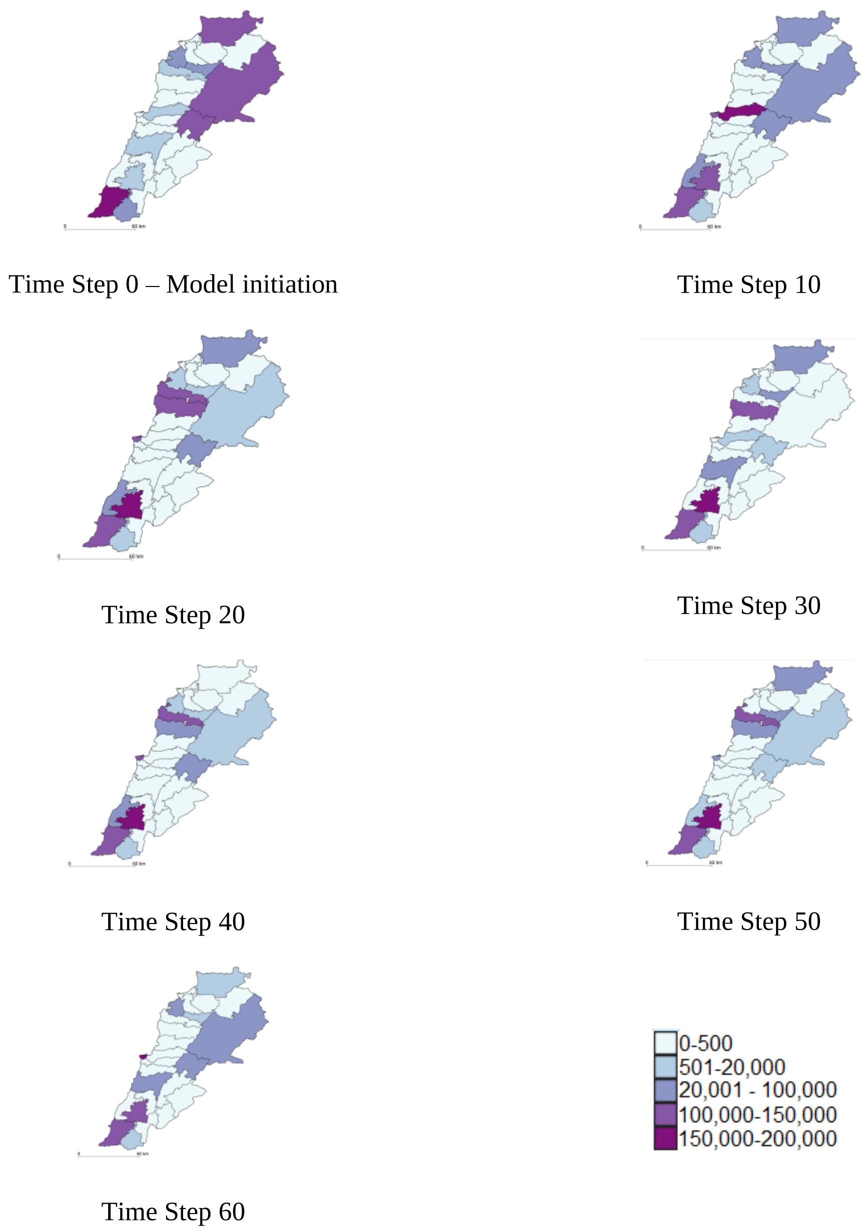


Figure 18 Spatial distribution of refugee agents every ten time steps within the Lebanon case study.

Jordan

In Jordan, the same Administrative Level was used which resulted in a location graph with 52 nodes, as depicted in Figure 19.

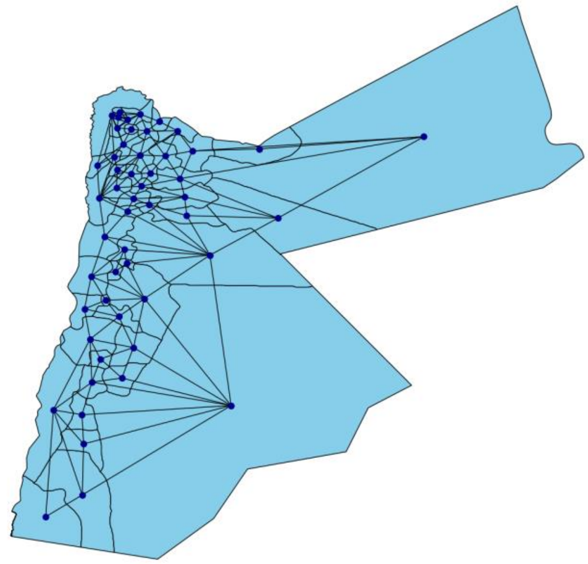


Figure 19 Location graph of Lebanon with 30 location nodes.

There were 21,984 refugee agents at initialization which grew to 712,400 refugees in Condition 1 (with social networks) and 789,088 in Condition 2 (without social networks). The actual number of refugees used to validate the simulation was 702,970 (UNHCR, 2020h) – this number is lower than the simulation numbers because approximately 10,000 refugees had returned to Syria during the study window. Model error (NMAE) is 0.19 for Condition 1 with social networks and 0.21 for Condition 2 without social networks, per Figure 20.

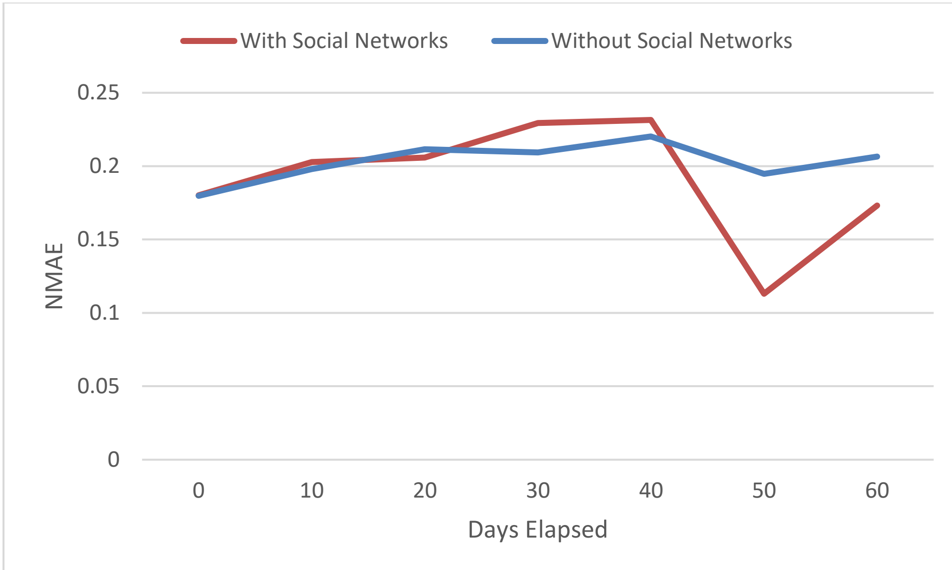


Figure 20 Error over time for the Jordan case study. Model error is 0.19 for Condition 1 with social networks and 0.21 for Condition 2 without social networks.

The spatial distribution of refugee movement throughout the simulation in Condition 1 is available in Figure 21.

Figure 21 Spatial distribution of refugee agents every ten time steps within the Jordan case study.

5.6 Discussion

The results presented here across the primary case study and two additional case studies, each run with two distinct test conditions, confirm that the inclusion of social networks serves to reduce model error across all study areas, as depicted in Table 12.

Table 12 Table comparing strengths of each model alongside their equivalents in the other model. Bolded elements indicated strengths.

	Model error with social networks	Model error without social networks	Simulated number of refugee agents	Actual number of refugee agents
Turkey	0.07	0.11	3,615,387	4,074,693
Lebanon	0.15	0.15	1,294,855	944,613
Jordan	0.19	0.21	712,400	702,970

Model error in Lebanon is also slightly lower with the social network condition. These results are presented in response to RQ1 affirming that the inclusion of social networks in agent-based simulations of forced migration positively affects the accuracy of bespoke models. In the case of Turkey, the model underestimated the number of refugees in the country as the rate of Syrian refugees entering the country during the study window was highly variable each day. In the cases of Jordan and Lebanon, during the study window, only a few refugees were entering the country while many were returning to their home state of Syria. For this reason, the model overestimates the number of refugee agents in the model because the model does not allow for refugee agents to leave the simulation environment once they have entered, thereby preventing the modeling of returnees. This topic is discussed in the following section.

Considering that the inclusion of social networks reduces model error and the state-of-the-art FLEE model does not include social networks, this is considered the primary methodological contribution of this model. There are several additional aspects of this model that can be compared to the FLEE model. Figure 22 presents a comprehensive annotation of the strengths and weaknesses of each model.

	My Model	FLEE Model
Environment	• Simulation environment creation automated • Applicable to any spatial extent	• Simulation environment created manually • Applicable only to a spatial extent containing pre-existing refugee aid infrastructure
Model	• Considers social networks • Scalable; tested up to 25M agents • Agents can be seeded anywhere • Transit distance not included in model logic	• No plans to consider social networks • Scalable; tested up to 1M agents • Agents must be seeded at conflict zones • Transit distance included in model logic
Simulation Package	• Stored on personal public GitHub • Tested in 3 AOIs • Model output in shapefile format • Model statistical output requires further interpretation	• Mature, releasable simulation framework • Tested in 5 AOIs • Model output is purely statistical • Model statistical output is complete

Figure 22 Comparison chart of strengths and weaknesses of this model and the state-of-the-art FLEE model.

Simulation Environment

There are several notable design differences in the simulation environments of the two models. One very important difference is that the FLEE model addresses migration to and from known, official refugee camps. This is the framework’s single greatest deficiency for two reasons: 1) migration is modeled in a constrained area represented by official refugee camps and migration to any other location in an AOI is not possible, and

2) given the model's need to initiate using official refugee camps, the model is not useful at the beginning of a refugee crisis when official camps have not yet been established. As such, the reporting of results is largely based on the population of refugees in camps as opposed to the spatial distribution of those refugees throughout a country. The FLEE model tackles a much more bounded analytic challenge than that addressed in this model. The model proposed here predicts refugee population movement throughout a county notwithstanding existing or planned refugee camp locations. This is the first improvement this model offers to the FLEE model. This model allows refugees to move more organically across geographic terrain by using a geospatial network base layer that is not confined to regions with hostile or friendly camp locations. The geospatial network in this model is of higher spatial fidelity than that of the FLEE model and yet is still flexible enough to replicate the spatial extent of the FLEE model easily. Furthermore, the FLEE model requires a highly manual analytic process to create the simulation environment which takes several hours even using a relatively constrained spatial extent. In contrast, this model creates a simulation environment from input shapefiles which are easily available in open sources in less than 20 seconds using a data pre-processing script.

Additionally, in the FLEE model, refugees are seeded in the simulation at conflict zones. In the case of Turkey, the origin of refugee flows is a collection of official and unofficial border crossings, which are included in this model as seed locations so the FLEE model's method of seeding is inadequate. In other locations, such as CAR or DRC, where refugees are likely seeded from conflict zones, not border crossings, the FLEE model's method may be more appropriate, but still not extensible to other geographies where this is not the case. Alternatively, this model specifies refugee seed locations using a global model parameter with string input, such that an analyst or researcher could specify the district nodes where refugees will seed regardless of whether they are district nodes that contain border crossings, district nodes that contain conflict events, both, or neither.

Social Networks

In comparing the two models, they roughly contain the same set of input parameters, despite the fact that these input parameters influence agent movement across different geographic areas. These parameters address variables such as existing refugee population, areas of conflict or threat, and areas of asylum, e.g. known refugee camps, with the notable exception that the FLEE model includes transit distance as a factor in the model where this model does not. The variable that is of paramount importance in a refugee's decision-making process, however, is the influence of the refugee's social network on deciding where to go. This can be ascribed to strong or weak ties who have sent direct messages to a refugee indicating a destination location. This variable can also be tied to social network groups or channels that represent unofficial collections of refugees or unofficial aid organizations, both of which indicate destination locations where groups of unregistered refugees reside that are not prior acquaintances of the migrant. The FLEE model does not include such a variable. Given that the comparison of the two model conditions of this model revealed that error was reduced when social networks were included in agent decision logic, perhaps this is the reason for the lower model error found in this model in comparison with the FLEE model.

Model Development, Testing, and Use

In terms of model accuracy, a primary deficiency of the FLEE model is in the variability of its accuracy in the first 30-90 days of the simulation. In essence, the FLEE model's performance is best over one year from the first day of the simulation, which is not as relevant in operational contexts where aid organizations may be attempting to decide where to place camps at the beginning of a conflict. In terms of visualization, the model presented in this dissertation outputs statistics and shapefiles. In dealing with shapefiles both for input and output purposes, resources are available in open sources without preliminary analysis and visualization is possible without further conversions when the simulation is complete. That said, the statistics that the FLEE model outputs are complete where the statistics that this model outputs require further interpretation. This does, however, allow an analyst or a researcher to calculate his or her desired error statistic instead of relying on that calculated endogenous to the model itself, e.g., NMAE, Mean Relative Error (MRE), Mean Root Square Error (MRSE), Mean Absolute Scaled Error (MASE), etc. Finally, although the results reported for this model are encouraging, the FLEE model is still a more mature, fully released ABM framework tested on 5+ AOIs. Subsequent sections will address the suggestions for maturation of this work.

6. CONCLUSION AND FUTURE WORK

In Chapter 1, several research questions and contributions to migration studies, sociology, and CSS communities were put forth. Chapter 2 reviewed relevant background materials and scholarly literature to contextualize the research. The methodology discussion in Chapter 3, the discussion of computation in Chapter 4, and the case study presented in Chapter 5 contain substantive contributions to migration studies and CSS. In this chapter, those contributions are reviewed, and the initial research questions set forth in Chapter 1 are revisited. Section 6.1 addresses the contributions in the context of the research questions and Section 6.2 concludes this dissertation with a discussion of future work and how the research detailed herein can be further matured within the CSS and social science communities.

6.1 Summary of Dissertation Results

This dissertation has diligently explored three research questions, the contributions of which apply broadly to the social science and CSS communities. Below is a review of those research questions and the outcomes they contribute to forced migration studies.

Research Question 1

Research question 1 is steeped in the emerging theory within forced migration studies concerning migrant social networks. *RQ1 To what extent does the consideration of social networks in forced migration models improve model accuracy?* It has been shown in Chapter 2 why social networks are of paramount importance to forced migration modeling efforts, yet it was also brought to light that no large-scale empirical ABM of forced migration to date has included an explicit refugee social network component. Preliminary theoretical modeling efforts around forced migration has shown that these social networks are both important and influential in forced migration modeling, and the outcomes from RQ1 take a step towards filling the research gap by designing and developing an ABM that considers migrant social networks in the modeling mechanics. The primary contribution put forth in this research is the first empirical ABM of forced migration to consider migrant social networks. In so doing, it was found that the inclusion of social networks improves model accuracy as implemented in this model. Substantially more research into mechanisms for the inclusion of social networks in forced migration ABMs is required before these models are tested thoroughly enough to be applied by aid organizations. This research is discussed further in Section 5.2.

To the theory of forced migration, the implications of these results substantiate the growing body of knowledge regarding the use of the SNT in forced migration modeling discussed in Hinsch and Bijak (2019), Reinhardt et al. (2019), Al-Khulaidy and Swartz (2020), Blumenstock, Chi, & Tan (2019), and Collins and Frydenlund (2016) and, for the first time, applies this theory to a real-world forced migration scenario at scale. The computational methods with which SNT has been applied are available for replication, maturation, and re-application to other forced migration scenarios in the future, constituting a methodological contribution. Finally, in practical applications of the model, the results in Chapter 4 demonstrate that including these social networks reduces model error. Most importantly, however, the results of this application further support this vein of analytic inquiry by showing that refugees do, indeed, make different decisions and settle in different locations as a result of forced displacement when being influenced by social networks than when not.

Research Question 2

Research question 2 is intended to speak to the CSS modeler and researcher who aspires to create ABMs of his or her own. *How can ABMs of forced migration be designed and developed at scale to facilitate further investigation?* One of the primary contributions of this dissertation is the computational methodologies provided to assist a researcher in developing a simulation quickly, specifically concerning the simulation environment. While the FLEE model does provide a generalizable framework and an open-source codebase for creating a simulation in a new AOI, the process is manual, labor-intensive, and error prone, requiring hand calculations and use of open-source GIS and navigation technologies. The methodologies presented in Chapter 3 include a codebase separate from the full simulation code exclusively for the creation of a network-based simulation environment in any AOI regardless of pre-existing refugee crisis response infrastructure or aid organization presence. This code is available at https://github.com/richey17/mig and is stored in the file pre_process_basic.py with associated README. Chapter 3 demonstrates how this code can be used to instantiate a new simulation environment in any AOI with varying levels of modeling granularity and control using different administrative levels in input shapefiles. This satisfies the aspect of the research question around how this research contributes to scalable simulation design and development. The second part of the question is how this particular simulation can be deployed at scale given that a 60-day simulation run with 4M refugee agents takes 96+ hours to run on standard consumer computing hardware. In response to this aspect of the question, the codebase is fully parallelized and designed to run in a Linux environment reducing full model runtime to just 10-13 hours. This makes testing, validation, and experimentation possible at scale – 2M-10M refugee agents. The computational experiments reported in Section 4.4. further report testing with up to 25M agents.

Research Question 3

Research question 3 addresses the application and validation of the model designed and developed under RQ1 to a prescient case study – the Syria/Turkey forced migration event. *How can the model(s) developed in response to RQ1 and RQ2 be applied to predict where are refugee populations are likely to move during and directly following a forced migration event in the context of a case study?* Confidently, from the case study presented in Chapter 4, it is shown that ABMs are, indeed, capable of empirically modeling and predicting forced migration flows at least insofar as the Syrian migration into Turkey is concerned. Model error is comparatively low (0.07 NMAE to the FLEE model's 0.33 NMAE in the same AOI) and several structural and design improvements have been made to the state-of-the-art competitor model. Notably, the full spatial extent of the AOI in question is modeled here where the FLEE model is only capable of modeling a fraction of the country of Turkey. The implication here is that this will allow for a more organic modeling effort and for modeling the emergence of unofficial refugee communities within the country – something the FLEE model is simply not designed to accomplish. With 98 percent of refugees in Turkey residing outside of camps, the ability to model the emergence of these unofficial communities is paramount to understanding the refugee crisis in Turkey. Additionally, model error is stable over time indicating that this model could reasonably be applied at the onset of a new Syrian refugee crisis and used for predictions 30-90 days in the future. These methodological improvements are a primary contribution of this research to the CSS community.

In summary, the research developed in the answering of these research questions has provided two primary contributions to the scientific and modeling community:

1. the first scalable ABM of forced migration to consider migrant social networks and the substantiation that the inclusion of these networks does, indeed, influence migrant decision-making in real-world conditions
2. methodological improvements to the state-of-the-art FLEE model which contribute to the extensibility and applicability of the model to include publicly available computational tools and methodologies to facilitate replication of this model and other modeling efforts in the CSS community, suggestions for which are discussed in the following section.

6.2 Future Work

It is my hope that the data pre-processing pipelines created in response to RQ2 will be used by other researchers to create simulation environments in a variety of AOIs quickly. This automation will simplify a large task for CSS modelers in the future who may wish to pursue the development of an authoritative ABM framework for forced migration, making possible the creation of many simulation environments globally for testing and validation purposes. For this reason, these codebases are available publicly at https://github.com/mrichey17/mig to facilitate further development. The simulation environment codebase is bereft of any model logic (pre_process_basic.py and pre_process_conversion.py), so new model logic can be applied to that simulation infrastructure or the model logic presented in this dissertation matured and applied to other AOIs using the full ABM codebase (run_abm_shared_mem.py).

The work presented herein necessitates application to other AOIs to assess its generalizability to other forced migration phenomena which would entail the possible addition of model design features and mechanics, such as additional agent classes. It is recommended to start with other AOIs for which historical data exist such as CAR and Colombia, though forced migration models for Eastern and parts of Western Europe would be prescient given current events. Regarding this specific model, the work of Reinhardt et al. (2019) sets a brilliant example of developing ABMs in pairs in different programming languages. This model could be implemented using another programming language for comparative purposes, or to facilitate greater scalability through MASON or another ABM parallelization or distributed computing framework.

The most important area of further research, however, is that of the social influence component. While this dissertation has contributed greatly to that body of knowledge by creating the first ever ABM of forced migration to explicitly include social networks, there is still only one. Future research should develop and test a variety of models around the implementation of social network and social media mechanics and build upon the model logic implemented here with those additions. There are several categories of social network dynamics within which several models could be developed for each to assess the impact of social influence on forced migration through a variety of lenses (Frydenlund & De Kock, 2020). Specifically, there are several social influence mechanics in voluntary migration models or that stem from the sociology literature that would be applicable here. The social network models recommended for future investigation are described in Table 14.

Table 13 Five suggested models for future work around social network and social influence mechanics

Suggested Social Network Models	Description	Reference
Dynamic social networks	Social ties become stronger or weaker during the simulation depending on agent co-location or separation. Social ties can also be created.	Reichlova, 2005
Strong and weak ties	Two kinds of social ties (strong and weak) are implemented in the model and agents' reliance on each is assessed. Empirical literature favors strong ties in migration research but recent developments in forced migration research indicate weak ties may play a stronger role.	Bakshy et al., 2012; Centola and Macey, 2007; Granovetter, 1977
Multiple agent classes	Agent classes with different model logic and priorities depending on length of time in the simulation. Agent classes with different model logic and priorities depending on migration intentions.	Robinson and Segrott, 2002
Explicit information exchange	Develop more sophisticated information exchange mechanisms for agents to communicate in a more detailed way about the forced migration environment.	Hinsch and Bijak, 2019; Reinhardt et al., 2019
Structural assessments	Analyze structures and substructures of migrant social networks to determine which types of social networks and what structural features of those networks are most realistic for modeling forced migrant social ties.	Blumenstock, Chi, and Tan, 2019; Garip, 2008

First, and likely the most computationally intensive of them all, is the dynamic evolution of social networks during model runs. Reichlova (2005) uses agents in a model of voluntary migration that have a strong preference for being in the same location as their social ties. These social ties also become stronger with every step that agents are co-located and, conversely, degrade with spatial separation of agents. This model logic is preliminarily included in the full codebase of this dissertation but was not implemented or tested fully. Along these same lines, this model could also allow for the creation of new social ties during simulation runtime; for example, if two agents are co-located in a refugee camp for more than 14 steps, a social tie would be created, and then strengthen or decay based on continued co-location or separation. This model would likely add realism to the model and it would be interesting to see if such a mechanic decreased overall model error.

The second model should explore the influence of both strong and weak ties on agent decision-making and model error separately. When considering social networks in migration, intuition suggests that strong ties, or familial ties, are what impacts migrant decision-making the most. In voluntary migration this is likely true; family members migrating to locations where other family members have expatriated (Hanson, 2006). The empirical literature follows this intuition and emphasizes stronger ties (Simon, 2019; King, 2012). Research specifically into refugees' use of social media, however, has shown that weak ties are just as important, if not more important, in forced migration contexts; for example, refugees exchanging phone numbers with unknown people to get live updates on a migration route from a group of people that departed a few hours prior (Dekker et al., 2018; Borkert, Fisher, & Yafi, 2018; Richey, 2014b). It is therefore relevant for a social networking model to study the effects of strong ties versus weak ties in agent decision-making, implementing these mechanics within a model in keeping with Grannovetter's (1977) original theory (Bashy et al., 2012). This model could be implemented by expanding the model logic of this dissertation model around friendship and kinship ties, which are distinguished as model parameters but do not lead to different agent logic or behaviors.

A third model could propose multiple agent classes as a way to test different agent decision-making logic depending upon an agent's progress through the migration journey. DeAngelis and Diaz (2019) review many implementations of evolutionary decision logic that could be applicable here, to include ML-based decision logic. Multiple agent classes are initially suggested to include an agent class comprised of refugees already in the target country at the time of simulation start and a second agent class comprised of refugees that have recently crossed a border or are seeded during the course of the simulation. Refugees in the first agent class would be more inclined to stay at their present location, provided their present location was sufficient in all other variables. These agents would serve as pull factors in the social network, attracting other agents to their present location. Their logic should be biased more strongly to remain in their present locations provided they are located in an urban center or otherwise outside an official refugee camp. The second agent class would require logic that is much more responsive to pull factors; specifically, the social network. In short, movement of one refugee class needs to be less dependent on the social network than the movement of another refugee class. Multiple agent classes could also be explored to represent refugees with different goals, e.g., to return to the country of origin, to settle in the country of asylum, or to transit to another country. It would be difficult, however, to empirically justify how many of each type of refugee to initiate in the model as many refugees are undecided at time of travel (Robinson & Segrott, 2002; Baban, Ilcan, & Rygiel, 2017; İçduygu, 2015; International Crisis Group, 2018; Richey, 2014b).

The recent work of Hinsch and Bijak (2019) and Reinhardt et al. (2019) have explored the nature of information exchange and social capital in the context of a highly abstract migration ABM. This application presents ways in which agents can exchange information about their current statuses and the environment in much more detail than the naïve social networking agents implemented in this model. A new model should implement information exchange, social, influence, and social capital in a much more rigorous and explicit manner to evolve the sophistication of agent-to-agent communication regarding aspects of forced migration.

The fifth and final suggested model is a model that explores the impact structural elements of the social network may have on agent decision-making and model error. The model presented in this dissertation creates agent social networks randomly without conforming to any known network creation logic such as preferential attachment or the deliberate creation of small world networks. No effort was put, in this dissertation, towards analyzing the structures and substructures of agent social networks and the impacts these may have on the model, e.g., average path length, network density, network transitivity, clustering, etc. Research around social networks is substantive and it is not lacking for theories of social influence and contagion and it is proposed that they be explored in the context of their impact on forced migrant social networks (Blumenstock, Chi, & Tan, 2019; Garip, 2008; McAuliffe, 2013; Herrera, Armelini, & Salvaj, 2015).

The work presented in this dissertation represents meaningful steps in the direction of considering a social networking component in forced migration modeling. It is the hope that this line of research will be continued leveraging many of the model design mechanics and computational methods outlined in the previous chapters, if not to the benefit of the CSS modelers, sociologists, and aid organizations charged with addressing refugee crises worldwide, then to the benefit of the refugees themselves who, in times of instability and insecurity, seek asylum and safety above all else.

REFERENCES

Aaby, B. G., Perumalla, K. S., & Seal, S. K. (2010, March). Efficient simulation of agent-based models on multi-GPU and multi-core clusters. In *Proceedings of the 3rd international icst conference on simulation tools and techniques* (pp. 1-10).

Adhikari, P. (2013). Conflict-Induced Displacement, Understanding the Causes of Flight. *American Journal of Political Science*, 57(1), 82-89.

Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In *Action control* (pp. 11-39). Springer, Berlin, Heidelberg.

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, *50*(2), 179-211.

Alhanaee, A., & Csala, D. (2015). *Motives to flee: Modeling Syrian refugee crisis*. Working paper.[Online] Available at https://www. researchgate. net/publication/299049673 [Accessed Apr 28, 2017] DOI: 10.13140/RG. 2.1. 2986.1524.

Al-Khulaidy, A., & Swartz, M. (2020, May). Along the border: an agent-based model of migration along the United States-Mexico border. In *2020 Spring Simulation Conference (SpringSim)* (pp. 1-12). IEEE.

Andersen, R., Chung, F., & Lang, K. (2006, October). Local graph partitioning using pagerank vectors. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)* (pp. 475-486). IEEE.

Andreev, K., & Racke, H. (2006). Balanced graph partitioning. *Theory of Computing Systems*, *39*(6), 929-939.

Aral, S., Muchnik, L., & Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, 106(51), 21544-21549.

Arango, J. (2000). Explaining migration: a critical view. *International social science journal*, *52*(165), 283-296.

Arentze, T. A., Kowald, M., & Axhausen, K. W. (2013). An agent-based random-utility-maximization model to generate social networks with transitivity in geographic space. *Social Networks, 35*(3), 451-459.

Azose, J. J., & Raftery, A. E. (2019). Estimation of emigration, return migration, and transit migration between all pairs of countries. *Proceedings of the National Academy of Sciences, 116*(1), 116-122.

Baban, F., Ilcan, S., & Rygiel, K. (2017). Syrian refugees in Turkey: Pathways to precarity, differential inclusion, and negotiated citizenship rights. *Journal of Ethnic and Migration Studies*, *43*(1), 41-57.

Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012, April). The role of social networks in information diffusion. In *Proceedings of the 21st international conference on World Wide Web* (pp. 519-528).

Ballas, D., Clarke, G. P., & Wiemers, E. (2005). Building a dynamic spatial microsimulation model for Ireland. *Population, Space and Place*, *11*(3), 157-172.

Barbosa Filho, H. S., de Lima Neto, F. B., & Fusco, W. (2011, April). Migration and social networks—an explanatory multi-evolutionary agent-based model. In *2011 IEEE Symposium on Intelligent Agent (IA)* (pp. 1-7). IEEE.

Barthel, F., & Neumayer, E. (2015). Spatial dependence in asylum migration. *Journal of Ethnic and Migration Studies*, *41*(7), 1131-1151.

Batty, M. (2008). Fifty years of urban modeling: Macro-statics to micro-dynamics. In *The dynamics of complex urban systems* (pp. 1-20). Physica-Verlag HD.

Bayar, M., & Aral, M. M. (2019). An analysis of large-scale forced migration in Africa. *International journal of environmental research and public health*, *16*(21), 4210.

Benattia, T., Armitano, F., & Robinson, H. (2015). Irregular Migration Between West Africa, North Africa and the Mediterranean. *Paris: Altai Consulting*.

Benenson, I., & Torrens, P. (2004). *Geosimulation: Automata-based modeling of urban phenomena*. John Wiley & Sons.

Benito-Ostolaza, J. M., Hernández, P., Palacios-Marqués, D., & Vila, J. (2015). Modeling local social migrations: A cellular automata approach. *Cybernetics and Systems*, *46*(3-4), 287-302.

Bergstrand, J. H. (1985). The gravity equation in international trade: some microeconomic foundations and empirical evidence. The review of economics and statistics, 474-481.

Bhattacharya, P., Ekanayake, S., Kuhlman, C. J., Lebiere, C., Morrison, D., Swarup, S., ... & Orr, M. G. (2019, May). The Matrix: An Agent-Based Modeling Framework for Data Intensive Simulations. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems* (pp. 1635-1643). International Foundation for Autonomous Agents and Multiagent Systems.

Bijak, J. (2006, April). Forecasting international migration: Selected theories, models, and methods. Warsaw, Poland: Central European Forum for Migration Research.

Biondo, A. E., Pluchino, A., & Rapisarda, A. (2012). Return migration after brain drain: A simulation approach. *arXiv preprint arXiv:1206.4280*.

Birkin, M., Malleson, N., Hudson-Smith, A., Gray, S., & Milton, R. (2011). Calibration of a spatial simulation model with volunteered geographical information. *International Journal of Geographical Information Science*, *25*(8), 1221-1239.

Blanchet, K., Fouad, F. M., & Pherali, T. (2016). Syrian refugees in Lebanon: the search for universal health coverage. *Conflict and health*, *10*(1), 1-5.

Blumenstock, J. E., Chi, G., & Tan, X. (2019). Migration and the value of social networks.

Bohra-Mishra, P., & Massey, D. S. (2011). Individual decisions to migrate during civil conflict. *Demography*, *48*(2), 401-424.

Borgatti, S. P., & Ofem, B. (2010). Social network theory and analysis. *Social network theory and educational change*, 17-29.

Borkert, M., Fisher, K. E., & Yafi, E. (2018). The best, the worst, and the hardest to find: How people, mobiles, and social media connect migrants in (to) Europe. *Social Media+ Society*, *4*(1), 2056305118764428.

Bretagnolle, A., & Pumain, D. (2010). Simulating urban networks through multiscalar space-time dynamics: Europe and the united states, 17th-20th centuries. *Urban Studies*, *47*(13), 2819-2839.

Bura, S., Guérin-Pace, F., Mathian, H., Pumain, D., & Sanders, L. (1996). Multiagent systems and the dynamics of a settlement system. *Geographical analysis*, *28*(2), 161-178.

Calvo, G. A. (1978). Urban unemployment and wage determination in LDC's: Trade unions in the Harris-Todaro model. *International Economic Review*, 65-81.

Cassarino, J. P. (2004). Theorising return migration: The conceptual approach to return migrants revisited. *International Journal on Multicultural Societies (IJMS)*, *6*(2), 253-279.

Castles, S. (2003). Towards a sociology of forced migration and social transformation. *Sociology*, 37(1), 13-34.

Centola, D., & Macy, M. (2007). Complex contagions and the weakness of long ties. *American journal of Sociology*, *113*(3), 702-734.

Cherri, Z., González, P. A., & Delgado, R. C. (2016). The Lebanese–Syrian crisis: impact of influx of Syrian refugees to an already weak state. *Risk management and healthcare policy*, *9*, 165.

Chun, Y., & Griffith, D. A. (2011). Modeling network autocorrelation in space–time migration flow data: an eigenvector spatial filtering approach. *Annals of the Association of American Geographers*, *101*(3), 523-536.

Cioffi-Revilla, C. (2005). A canonical theory of origins and development of social complexity. Journal of Mathematical Sociology, 29(2), 133-153.

Cioffi-Revilla, C., Rogers, J.D., Schopf, P.S., Luke, S., Bassett, J.C., Hailegiorgis, A., Kennedy, W.M., Froncek, P., Mulkerin, M., Shaffer, M., & Wei, E. (2015). MASON NorthLands: A Geospatial Agent-Based Model of Coupled Human-Artificial-Natural Systems in Boreal and Arctic Regions. *European Social Simulation Association (ESSA).*

Cohen, J. E., Roig, M., Reuman, D. C., & 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.

Collier, N., & North, M. (2012). Repast HPC: A platform for large-scale agent-based modeling. *Large-Scale Computing*, 81-109.

Collier, N., Ozik, J., & Macal, C. M. (2015, August). Large-scale agent-based modeling with repast HPC: A case study in parallelizing an agent-based model. In *European Conference on Parallel Processing* (pp. 454-465). Springer, Cham.

Collins, A. J., & Frydenlund, E. (2016, December). Agent-based modeling and strategic group formation: a refugee case study. In *2016 Winter Simulation Conference (WSC)* (pp. 1289-1300). IEEE.

Crawley, H. (2010). Chance or Choice?: Understanding why asylum seekers come to the UK. Refugee Council.

Crooks, A., Castle, C., & Batty, M. (2008). Key challenges in agent-based modelling for geo-spatial simulation. *Computers, Environment and Urban Systems,* 32(6), 417-430.

Curry, T., Croitoru, A., Crooks, A., & Stefanidis, A. (2019). Exodus 2.0: crowdsourcing geographical and social trails of mass migration. *Journal of Geographical Systems*, 21(1), 161-187.

Cushman, S. A., McKelvey, K. S., Hayden, J., & Schwartz, M. K. (2006). Gene flow in complex landscapes: testing multiple hypotheses with causal modeling. *The American Naturalist*, *168*(4), 486-499.

Cushman, S. A., Chase, M., & Griffin, C. (2010). Mapping landscape resistance to identify corridors and barriers for elephant movement in southern Africa. In Spatial complexity, informatics, and wildlife conservation (pp. 349-367). Springer, Tokyo.

Dabbaghian, V., Jackson, P., Spicer, V., & Wuschke, K. (2010). A cellular automata model on residential migration in response to neighborhood social dynamics. *Mathematical and Computer Modelling*, *52*(9-10), 1752-1762.

DeAngelis, D. L., & Diaz, S. G. (2019). Decision-making in agent-based modeling: A current review and future prospectus. *Frontiers in Ecology and Evolution*, *6*, 237.

Dekker, R., Engbersen, G., Klaver, J., & Vonk, H. (2018). Smart refugees: How Syrian asylum migrants use social media information in migration decision-making. *Social Media+ Society*, *4*(1), 2056305118764439.

Dekkers, G. (2015). *On the modelling of immigration and emigration using LIAM2* (pp. 1-13, Rep.). Federaal Planbureau: Economische analyses en vooruitzichten.

De Haas, H. (2010). Migration and development: A theoretical perspective. *International migration review*, *44*(1), 227-264.

De Koch, C. (2019). *A Framework For Modelling Conflict-Induced Forced Migration According To An Agent-Based Approach* (Doctoral dissertation).

Ding, C. H., He, X., Zha, H., Gu, M., & Simon, H. D. (2001, November). A min-max cut algorithm for graph partitioning and data clustering. In *Proceedings 2001 IEEE international conference on data mining* (pp. 107-114). IEEE.

Disney, G., Wiśniowski, A., Forster, J. J., Smith, P. W., & Bijak, J. (2015). Evaluation of existing migration forecasting methods and models. *Report for the Migration*

Advisory Committee: Commissioned research. ESRC Centre for Population Change, University of Southampton.

Docquier, F., Peri, G., & Ruyssen, I. (2014). The cross-country determinants of potential and actual migration. *International Migration Review*, *48*(1_suppl), 37-99.

Dorigo, G., & Tobler, W. (1983). Push-pull migration laws. *Annals of the Association of American Geographers*, *73*(1), 1-17.

Durand, J., & Massey, D. S. (Eds.). (2004). *Crossing the border: Research from the Mexican migration project*. Russell Sage Foundation.

Echevarria-Coco, J., & Gardeazabal, J. (2020). A Spatial Model of Internal Displacement and Forced Migration. *Journal of Conflict Resolution*, 0022002720958470.

Edmonds, B. (1999). Modelling bounded rationality in agent-based simulations using the evolution of mental models. In *Computational techniques for modelling learning in economics* (pp. 305-332). Springer, Boston, MA.

Edwards, S. (2008). Computational tools in predicting and assessing forced migration. *Journal of Refugee Studies*, *21*(3), 347-359.

Elffers, H., Reynald, D., Averdijk, M., Bernasco, W., & Block, R. (2008). Modelling crime flow between neighbourhoods in terms of distance and of intervening opportunities. *Crime Prevention and Community Safety*, *10*(2), 85-96.

Epstein, G. S. (2008). Herd and network effects in migration decision-making. Journal of Ethnic and Migration Studies, 34(4), 567-583.

Epstein, G. S., & Gang, I. N. (2006). The influence of others on migration plans. *Review of Development Economics*, *10*(4), 652-665.

Espíndola, A. L., Silveira, J. J., & Penna, T. J. P. (2006). A Harris-Todaro agent-based model to rural-urban migration. *Brazilian journal of physics*, *36*(3A), 603-609.

Fishbein, M. (1979). A theory of reasoned action: some applications and implications. *Nebraska Symposium on Motivation, 27*, 65–116.

Flowerdew, R., & Lovett, A. (1988). Fitting constrained Poisson regression models to interurban migration flows. *Geographical Analysis*, *20*(4), 297-307.

Fotheringham, A. S. (1981). Spatial structure and distance-decay parameters. *Annals of the Association of American Geographers*, *71*(3), 425-436.

Fotheringham, A. S. (1983). A new set of spatial-interaction models: the theory of competing destinations. *Environment and Planning A: Economy and Space*, *15*(1), 15-36.

Fotheringham, A. S., & O'Kelly, M. E. (1989). *Spatial interaction models: formulations and applications* (Vol. 1, p. 989). Dordrecht: Kluwer Academic Publishers.

Frydenlund, E. & De Kock, C. (2020). Agent-based modeling within forced migration research: A review and critique. *Refugee Review: Emerging Issues in Forced Migration*, *4*, 53-68.

Frydenlund, E., Foytik, P., Padilla, J. J., & Ouattara, A. (2018, December). Where are they headed next?: modeling emergent displaced camps in the DRC using agent-based models. In *Proceedings of the 2018 Winter Simulation Conference* (pp. 22-32). IEEE Press.

Frydenlund, E., Şener, M. Y., Gore, R., Boshuijzen-van Burken, C., Bozdag, E., & de Kock, C. (2019). Characterizing the Mobile Phone Use Patterns of Refugee-Hosting Provinces in Turkey. In *Guide to Mobile Data Analytics in Refugee Scenarios* (pp. 417-431). Springer, Cham.

García-Díaz, C., & Moreno-Monroy, A. I. (2012). Social influence, agent heterogeneity and the emergence of the urban informal sector. *Physica A: Statistical Mechanics and its Applications*, *391*(4), 1563-1574.

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

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision-making. *Annual review of psychology*, *62*, 451-482.

Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In *Simple heuristics that make us smart* (pp. 3-34). Oxford University Press.

Gilbert, N. (2002, October). Varieties of emergence. In *Agent 2002 Conference: Social agents: ecology, exchange, and evolution, Chicago* (pp. 11-12).

Gilbert, N. (2007). Computational social science: Agent-based social simulation.

Gilbert, N., & Troitzsch, K. (2005). *Simulation for the social scientist*. McGraw-Hill Education (UK).

Giulietti, C., Wahba, J., & Zenou, Y. (2018). Strong versus weak ties in migration. *European Economic Review*, *104*, 111-137.

Goodchild, M. F., & Smith, T. R. (1980). Intransitivity, the spatial interaction model, and US migration streams. *Environment and Planning A*, *12*(10), 1131-1144.

Granovetter, M. S. (1977). The strength of weak ties. In *Social networks* (pp. 347-367). Academic Press.

Gray, J., Hilton, J., & Bijak, J. (2017). Choosing the choice: Reflections on modelling decisions and behaviour in demographic agent-based models. *Population Studies*, *71*(sup1), 85-97.

Grimm et al. (2020). The ODD Protocol for Describing Agent-Based and Other Simulation Models: A Second Update to Improve Clarity, Replication, and Structural Realism. *Journal of Artificial Societies and Social Simulation*, *23*(2).

Grimm, V., Berger, U., DeAngelis, D. L., Polhill, J. G., Giske, J., & Railsback, S. F. (2010). The ODD protocol: a review and first update. *Ecological modelling*, *221*(23), 2760-2768.

Grimm, V., Polhill, G., & Touza, J. (2017). Documenting social simulation models: the ODD protocol as a standard. In *Simulating social complexity* (pp. 349-365). Springer, Cham.

Groen, D. (2018, June). Development of a multiscale simulation approach for forced migration. In *International Conference on Computational Science* (pp. 869-875). Springer, Cham.

Groen, D., Knap, J., Neumann, P., Suleimenova, D., Veen, L., & Leiter, K. (2019). Mastering the scales: a survey on the benefits of multiscale computing software. *Philosophical Transactions of the Royal Society A*, *377*(2142), 20180147.

Gulden, T., Harrison, J. F., & Crooks, A. T. (2011, October). Modeling cities and displacement through an agent-based spatial interaction model. In *The Computational Social Science Society of America Conference*.

Gurak, D. T., & Caces, F. (1992). Migration networks and the shaping of migration systems. *International migration systems: A global approach*, 150-176.

Gulyás 12, L., Szabó, A., Legéndi, R., Máhr, T., Bocsi, R., & Kampis, G. (2011). Tools for large scale (distributed) agent-based computational experiments. Presented at the *2011 CSSSA Annual Conference*, 10-12 October 2011, Santa Fe, New Mexico, USA.

Hanson, G. H. (2006). Illegal migration from Mexico to the United States. *Journal of Economic Literature*, *44*(4), 869-924.

Harris, J. R., & Todaro, M. P. (1970). Migration, unemployment and development: a two-sector analysis. *The American economic review*, *60*(1), 126-142.

Haug, S. (2008). Migration networks and migration decision-making. *Journal of ethnic and migration studies*, *34*(4), 585-605.

Hattle, A., Yang, K. S., & Zeng, S. (2016). Modeling the Syrian Refugee Crisis with Agents and Systems. UMAP Journal, 37(2).

Havinga, T., & Böcker, A. (1999). Country of asylum by choice or by chance: Asylum-seekers in Belgium, the Netherlands and the UK. *Journal of ethnic and migration studies*, *25*(1), 43-61.

Hébert, G. A., Perez, L., & Harati, S. (2018). An agent-based model to identify migration pathways of refugees: the case of Syria. In *Agent-Based Models and Complexity Science in the Age of Geospatial Big Data* (pp. 45-58). Springer, Cham.

Herrera, M., Armelini, G., & Salvaj, E. (2015). Understanding social contagion in adoption processes using dynamic social networks. *PloS one*, *10*(10), e0140891.

HDX. (2020). Humanitarian Data Exchange/Home/Datasets. Retrieved November 01, 2020, from https://data.humdata.org/dataset

Hendrickson, B., & Kolda, T. G. (2000). Graph partitioning models for parallel computing. *Parallel computing*, *26*(12), 1519-1534.

Hijazi, S., Lovatt, H., & Iraqi, A. (n.d.). Refugee Camps. Retrieved November 23, 2020, from https://ecfr.eu/special/mapping_palestinian_politics/refugee_camps/

Hinsch, M., & Bijak, J. (2019). Rumours lead to self-organized migration routes. In *The 2019 Conference on Artificial Life: How Can Artificial Life Help Solve Societal Challenges?*. Retrieved 2020, from https://eprints.soton.ac.uk/432965/1/Paper_ALife_2019.pdf

Hinsch, M., & Bijak, J. (2019). Rumours lead to self-organized migration routes.

Ibáñez, A. M., & Vélez, C. E. (2008). Civil conflict and forced migration: The micro determinants and welfare losses of displacement in Colombia. *World Development*, *36*(4), 659-676.

İçduygu, A. (2015). Syrian refugees in Turkey: The long road ahead. *Washington, DC: Migration Policy Institute*.

International Crisis Group. (2018). *Turkey's Syrian Refugees: Defusing Metropolitan Tensions* (248). https://d2071andvip0wj.cloudfront.net/248-turkey-s-syrian-refugees.pdf

Jagarnathsingh, A. (2019). *Lebanon's Border Regime: Fluid Rigidity, Foreign Interference, and Hybrid Security Assemblages* (pp. 1-59, Working paper No. 22). Beirut: Lebanon Suppor.

Karemera, D., Oguledo, V. I., & Davis, B. (2000). A gravity model analysis of international migration to North America. Applied Economics, 32(13), 1745-1755.

Kavak, H., Padilla, J. J., Lynch, C. J., & Diallo, S. Y. (2018, April). Big data, agents, and machine learning: towards a data-driven agent-based modeling approach. In *Proceedings of the Annual Simulation Symposium* (p. 12). Society for Computer Simulation International.

Kazil, J., & Verzemnieks, N. (2014). Mesa: Agent-based modeling in Python 3+. *URl: https://github. com/projectmesa/mesa*.

Kennedy, W. G. (2012). Modelling human behaviour in agent-based models. In *Agent-based models of geographical systems* (pp. 167-179). Springer, Dordrecht.

King, R. (2011). (working paper). *Theories and Typologies of Migration: an overview and a primer* (Ser. Willy Brandt Series of Working Papers in International Migration and Ethnic Relations, pp. 3–43). Malmo: Malmo University.

Kiran, M., Richmond, P., Holcombe, M., Chin, L. S., Worth, D., & Greenough, C. (2010, May). FLAME: simulating large populations of agents on parallel hardware architectures. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1* (pp. 1633-1636).

Klabunde, A. (2014). Computational economic modeling of migration. *Available at SSRN 2470525*.

Klabunde, A., & Willekens, F. (2016). Decision-making in agent-based models of migration: state of the art and challenges. *European Journal of Population*, *32*(1), 73-97.

Klabunde, A., Zinn, S., Willekens, F., & Leuchter, M. (2017). Multistate modelling extended by behavioural rules: An application to migration. *Population studies*, *71*(sup1), 51-67.

Kuschminder, K., De Bresser, J., & Siegel, M. (2015). Irregular migration routes to Europe and factors influencing migrants' destination choices. *Maastricht: Maastricht Graduate School of Governance*, 8-20.

Langley, S., Vanore, M., Siegel, M., Roosen, I., Rango, M., Leonardelli, I., & Laczko, F. (2016). The Push and Pull Factors of Asylum Related Migration: A Literature Review. European Asylum Support Office.

Łatek, M. M., Rizi, S. M. M., & Geller, A. (2013, December). Verification through calibration: an approach and a case study of a model of conflict in Syria. In *2013 Winter Simulations Conference (WSC)* (pp. 1649-1660). IEEE.

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

LeSage, J. P., & Fischer, M. M. (2010). Spatial econometric methods for modeling origin-destination flows. In *Handbook of applied spatial analysis* (pp. 409-433). Springer, Berlin, Heidelberg.

Lewer, J. J., & Van den Berg, H. (2008). A gravity model of immigration. *Economics letters*, 99(1), 164-167.

Li, L., Alderson, D., Doyle, J. C., & Willinger, W. (2005). Towards a theory of scale-free graphs: Definition, properties, and implications. *Internet Mathematics*, *2*(4), 431-523.

Lin, L., Carley, K. M., & Cheng, S. F. (2016, December). An agent-based approach to human migration movement. In *2016 Winter Simulation Conference (WSC)* (pp. 3510-3520). IEEE.

Luke, S., Cioffi-Revilla, C., Panait, L., Sullivan, K., & Balan, G. (2005). Mason: A multiagent simulation environment. *Simulation*, *81*(7), 517-527.

Maitland, C., & Xu, Y. (2015, March). A social informatics analysis of refugee mobile phone use: A case study of Za'atari Syrian refugee camp. TPRC.

Malakooti, A., Benattia, T., & Davin, E. (2013). Mixed migration: Libya at the crossroads. *report commissioned by the UNHRC-Tripoli November*.

Malakooti, A., & Davin, E. (2015). Migration Trends Across the Mediterranean: Connecting the Dots. *International Organization for Migration*.

Masad, D., & Kazil, J. (2015, July). MESA: an agent-based modeling framework. In *14th PYTHON in Science Conference* (pp. 53-60).

Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., & Taylor, J. E. (1993). Theories of international migration: A review and appraisal. *Population and development review*, 431-466.

Massey, D. S., & Zenteno, R. M. (1999). The dynamics of mass migration. *Proceedings of the National Academy of Sciences*, *96*(9), 5328-5335.

Mátyás, L. (1997). Proper econometric specification of the gravity model. *World Economy*, 20(3), 363-368.

McAuliffe, M. (2013). Seeking the views of irregular migrants: Decision-making, drivers and migration journeys. *A long way to go*, 103.

McKerns, M. M., Strand, L., Sullivan, T., Fang, A., & Aivazis, M. A. (2012). Building a framework for predictive science. *arXiv preprint arXiv:1202.1056*.

Miller, J. H. and Page, S. E. (2009), *Complex Adaptive Systems*, Princeton University Press, Princeton, NJ.

Moore, W. H., & Shellman, S. M. (2006). Refugee or internally displaced person? To where should one flee? *Comparative Political Studies*, 39(5), 599-622.

Moore, W. H., & Shellman, S. M. (2007). Whither will they go? A global study of refugees' destinations, 1965–1995. *International Studies Quarterly*, 51(4), 811-834.

Moreno, A., Rodríguez, J. J., Beltrán, D., Sikora, A., Jorba, J., & César, E. (2019). Designing a benchmark for the performance evaluation of agent-based simulation applications on HPC. *The Journal of Supercomputing*, *75*(3), 1524-1550.

Naudé, W. (2010a). The determinants of migration from Sub-Saharan African countries. *Journal of African Economies*, *19*(3), 330-356.

Naude, W. (2010b). Forced Migration from Sub-Saharan Africa: The Conflict–Environment Link. In *Environment, forced migration and social vulnerability* (pp. 43-55). Springer, Berlin, Heidelberg.

Nerbonne, J. (2010). Measuring the diffusion of linguistic change. Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1559), 3821-3828.

O'Donoghue, C., Redway, H., & Lennon, J. (2010). Simulating migration in the PENSIM2 dynamic microsimulation model. *International Journal of Microsimulation*, *3*(2), 65-79.

Padilla, J., Diallo, S., Barraco, A., Kavak, H., & Lynch, C. (2014). Cloud-based simulators: Making simulations accessible to non-experts and experts alike. *Proceedings - Winter Simulation Conference*. 2015. 10.1109/WSC.2014.7020192.

Parker, D. C., Entwisle, B., Rindfuss, R. R., Vanwey, L. K., Manson, S. M., Moran, E., ... & Mussavi Rizi, S. M. (2008). Case studies, cross-site comparisons, and the challenge of generalization: comparing agent-based models of land-use change in frontier regions. *Journal of Land Use Science*, 3(1), 41-72.

Parry, H. R., & Evans, A. J. (2008). A comparative analysis of parallel processing and super-individual methods for improving the computational performance of a large individual-based model. *Ecological Modelling*, *214*(2-4), 141-152.

Parry, H. R., & Bithell, M. (2012). Large scale agent-based modelling: A review and guidelines for model scaling. In *Agent-based models of geographical systems* (pp. 271-308). Springer, Dordrecht.

Perumalla, K. S., & Aaby, B. G. (2008, April). Data parallel execution challenges and runtime performance of agent simulations on GPUs. In *Proceedings of the 2008 Spring simulation multiconference* (pp. 116-123). Society for Computer Simulation International.

Poot, J., Alimi, O., Cameron, M., & Mare, D., The Gravity Model of Migration: The Successful Comeback of an Ageing Superstar in Regional Science. IZA Discussion Paper No. 10329, Available at SSRN: https://ssrn.com/abstract=2864830

Portugali, J., & Benenson, I. (1995). Artificial planning experience by means of a heuristic cell-space model: simulating international migration in the urban process. *Environment and Planning A*, *27*(10), 1647-1665.

Pumain, D., Sanders, L., Mathian, H., Guérin-Pace, F., & Bura, S. (1995). SIMPOP, A Multi-Agents Model for the Urban Transition. Sikos T. Bassa L. Fischer M. *Recent Developments in Spatial Information, Modeling and Processing*, Studies in Geography in Hungary, p. 13.

Quigley, C. (1979). *The evolution of civilizations: An introduction to historical analysis*. Liberty Fund.

Rae, A. (2009). From spatial interaction data to spatial interaction information? Geovisualisation and spatial structures of migration from the 2001 UK census. *Computers, Environment and Urban Systems, 33*(3), 161-178.

Raleigh, C., Linke, A., Hegre, H., & Karlsen, J. (2010). Introducing ACLED: an armed conflict location and event dataset: special data feature. *Journal of Peace Research*, 47(5), 651-660.

Rand, W. (2006, September). Machine learning meets agent-based modeling: when not to go to a bar. In *Conference on Social Agents: Results and Prospects*.

Ravenstein, E. G. (1885). The laws of migration. *Journal of the statistical society of London, 48*(2), 167-235.

Rehm, M. (2012). *Migration and remittances. An agent-based model* (Doctoral dissertation, New School University).

Reinhardt, O., Hilton, J., Warnke, T., Bijak, J., & Uhrmacher, A. M. (2018). Streamlining simulation experiments with agent-based models in demography. *Journal of Artificial Societies and Social Simulation*, *21*(3).

Reinhardt, O., Uhrmacher, A. M., Hinsch, M., & Bijak, J. (2019, December). Developing agent-based migration models in pairs. In *2019 Winter Simulation Conference (WSC)* (pp. 2713-2724). IEEE.

Reichlová, N. (2005). *Can the theory of motivation explain migration decisions?* (No. 97). Charles University Prague, Faculty of Social Sciences, Institute of Economic Studies.

Republic of Turkey Ministry of the Interior Directorate General of Migration Management Statistics. (2019, February 19). Retrieved May 04, 2019, from http://www.goc.gov.tr/icerik6/temporary-protection_915_1024_4748_icerik#

Richey, M. K. (2014a, March 03). What The Syrians Had To Say: Views from Istanbul (Part 5 Of A Multi-Part Series) [Web log post]. Retrieved November 03, 2020, from https://sourcesandmethods.blogspot.com/2014/03/what-syrians-had-to-say-views-from.html

Richey, M. K. (2014b). Geocultural Simulations: Populations and Predictions. In *Encyclopedia of US Intelligence-Two Volume Set* (pp. 1-13). Auerbach Publications.

Richey, M. K., & Mostowsky, Z. (2020). Leveraging HPC Techniques for Large-Scale Agent-Based Models in Python. In *Emerging Concepts and Innovation*

Technologies: Patterns of Life and Complexity, Virtual Interservice/Industry Training, Simulation, and Education (I/ITSEC) conference. Retrieved 2020, from https://www.viitsec.org/

Richmond, A. H. (1993). Reactive migration: Sociological perspectives on refugee movements. *Journal of Refugee Studies*, 6(1), 7-24.

Robinson, V., & Segrott, J. (2002). *Understanding the decision-making of asylum seekers* (Vol. 12). London: Home Office.

Rogers, A., Willekens, F., Little, J., & Raymer, J. (2002). Describing migration spatial structure. *Papers in Regional Science*, *81*(1), 29-48.

Roggio, B. (2013, September 29). Al Nusrah Front, Free Syrian Army seize border crossing to Jordan. Retrieved November 23, 2020, from https://www.longwarjournal.org/archives/2013/09/al_nusrah_front_free.php

Sarra, A. L., & Del Signore, M. (2010). A dynamic origin-constrained spatial interaction model applied to Poland's inter-provincial migration. *Spatial Economic Analysis*, *5*(1), 29-41.

Sawyer, R. K. (2001). Emergence in sociology: Contemporary philosophy of mind and some implications for sociological theory. *American Journal of Sociology*, *107*(3), 551-585.

Searle, C., & van Vuuren, J. H. (2021). Modelling forced migration: A framework for conflict-induced forced migration modelling according to an agent-based approach. *Computers, Environment and Urban Systems*, *85*, 101568.

Segawa, S., Kin, S., Kawamura, H., & Suzuki, K. (2015) Implementation of Massive Agent Model Using Repast HPC and GPU. *Systemics, Cybernetics, and Informatives*, *13*(2), 41-45.

Sewell, A. (2020, January 20). How Aid Groups Map Refugee Camps That Officially Don't Exist. Retrieved from https://www.wired.com/story/aid-groups-map-refugee-camps-officially-dont-exist/

Shackelford, B. B., Cronk, R., Behnke, N., Cooper, B., Tu, R., D'Souza, M., ... & Jaff, D. (2020). Environmental health in forced displacement: A systematic scoping review of the emergency phase. *Science of The Total Environment*, *714*, 136553.

Simini, F., González, M. C., Maritan, A., & Barabási, A. L. (2012). A universal model for mobility and migration patterns. *Nature*, *484*(7392), 96-100.

Simon, H. A. (1996), *The Sciences of the Artificial (3rd Edition)*, MIT Press, Cambridge, M. A.

Sokolowski, J. A., Banks, C. M., & Hayes, R. L. (2014, December). Modeling population displacement in the Syrian city of Aleppo. In *Proceedings of the Winter Simulation Conference 2014* (pp. 252-263). IEEE.

Stillwell, J. C. H. (1978). Interzonal migration: some historical tests of spatial-interaction models. *Environment and Planning A*, *10*(10), 1187-1200.

Stouffer, S. A. (1940). Intervening opportunities: a theory relating mobility and distance. *American Sociological Review*, 5(6), 845-867

Stouffer, S. A. (1960). Intervening opportunities and competing migrants. *Journal of regional science*, *2*(1), 1-26.

Sulaiman, S., Ali, U. I., & Hossen, M. (2019). A Predictive Model for the Population Growth of Refugees in Asia: A Multiple Linear Regression Approach. *Journal of Computational and Theoretical Nanoscience*, *16*(3), 1196-1202.

Suleimenova, D., Bell, D., & Groen, D. (2017). A generalized simulation development approach for predicting refugee destinations. *Scientific reports*, 7(1), 13377.

Suleimenova, D., & Groen, D. (2020). How policy decisions affect refugee journeys in South Sudan: a study using automated ensemble simulations. *Journal of Artificial Societies and Social Simulation*, *23*(1), 1-2.

Tobler, W. R. (1987). Experiments in migration mapping by computer. *The American Cartographer*, *14*(2), 155-163.

Tranos, E., Gheasi, M., & Nijkamp, P. (2015). International migration: a global complex network. *Environment and Planning B: Planning and Design*, 42(1), 4-22.

Trudgill, P. (1974). Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Language in Society*, 3(2), 215-246.

UNHCR. (2020a). Figures at a Glance. Retrieved November 03, 2020, from https://www.unhcr.org/en-us/figures-at-a-glance.html

UNHCR. (2020b, September). Turkey Fact Sheet. Retrieved November 01, 2020, from https://www.unhcr.org/tr/wp-content/uploads/sites/14/2020/10/UNHCR-Turkey-General-Fact-Sheet-September-2020-FINAL330.pdf

UNHCR. (2020c). Syria Emergency. Retrieved November 03, 2020, from https://www.unhcr.org/en-us/syria-emergency.html?query=syria

UNHCR. (2020d). Syria Regional Refugee Response. Retrieved November 05, 2020 from https://data2.unhcr.org/en/situations/syria#_ga=2.267211854.1534295988.1604547572-972670960.1604178683.

UNHCR (2020e). Emergencies. Retrieved November 05, 2020, from https://www.unhcr.org/en-us/emergencies.html

UNHCR. (2020f, August 13). Jordan Situation Map as of August 2020 (Lanscape). Retrieved November 23, 2020, from https://data2.unhcr.org/en/documents/details/78283

UNHCR. (2020g, September 30). Syria Regional Refugee Response: Lebanon. Retrieved November 23, 2020, from https://data2.unhcr.org/en/situations/syria/location/71

UNHCR. (2020h, November 04). Syria Regional Refugee Response: Jordan. Retrieved November 23, 2020, from http://data2.unhcr.org/en/situations/syria/location/36

UNOCHA. (2020, July 21). Turkey: Syria: Border Crossings Status. Retrieved November 01, 2020, from https://reliefweb.int/report/syrian-arab-republic/turkey-syria-border-crossings-status-21-july-2020-enartr

Vernon-Bido, D., Frydenlund, E., Padilla, J. J., & Earnest, D. C. (2017, April). Durable solutions and potential protraction: The syrian refugee case. In *Proceedings of the 50th Annual Simulation Symposium* (pp. 1-9).

Watts, J. (2016). Scale Dependency in Agent-Based Modeling: How Many Time Steps? How Many Simulations? How Many Agents?. In *Uncertainty and Sensitivity Analysis in Archaeological Computational Modeling* (pp. 91-111). Springer, Cham.

Wendel, S., & Dibble, C. (2007). Dynamic agent compression. *Journal of Artificial Societies and Social Simulation, 10*(2), 9.

Wilson, A. G. (1971). A family of spatial interaction models, and associated developments. *Environment and Planning A, 3*(1), 1-32.

Wittek, P., & Rubio-Campillo, X. (2012, December). Scalable agent-based modelling with cloud HPC resources for social simulations. In *4th IEEE International Conference on Cloud Computing Technology and Science Proceedings* (pp. 355-362). IEEE.

Wolfe, S. R., Sierhuis, M., & Jarvis, P. A. (2008, April). To BDI, or not to BDI: design choices in an agent-based traffic flow management simulation. In *SpringSim* (pp. 63-70).

Wozny, P. J. (2018). *A Value Sensitive Agent Based Simulation of the Refugee Crisis in the Netherlands* (Master's thesis).

Wu, Y., & Liu, S. (2014). A suggestion for computing objective function in model calibration. *Ecological informatics*, *24*, 107-111.

Xu, Y., Holzer, A., Maitland, c., & Gillet, D. (2017, November). Community building with co-located social media: a field experiment with Syrian refugees. In proceedings of the ninth international conference on information and communication technologies and development (p. 16). ACM.

Yano, K., Nakaya, T., & Ishikawa, Y. (2000). An analysis of inter-municipal migration flows in Japan using GIS and spatial interaction modeling. *Geographical review of Japan, Series B.*, *73*(2), 165-177.

Yano, K., Nakaya, T., Fotheringham, A. S., Openshaw, S., & Ishikawa, Y. (2003). A comparison of migration behaviour in Japan and Britain using spatial interaction models. *International Journal of Population Geography*, *9*(5), 419-431.

Yaylacı, F. G., & Karakuş, M. (2015). Perceptions and newspaper coverage of Syrian refugees in Turkey. *Migration Letters*, *12*(3), 238-250.

BIOGRAPHY

Melonie K Richey graduated from Mercyhurst University in 2014 with an MS in Intelligence Studies, an interdisciplinary analytic methods program focused on National Security and private industry intelligence work. During her time at Mercyhurst, she concentrated her graduate-level research in the domain of cognitive psychology; specifically, how to mitigate the effects of cognitive bias in analysis and decision-making. Secondarily, she focused a second avenue of independent research on the modeling of forced migration flows in various geographic locations worldwide. Ms. Richey also graduated from the University of Florida in 2011 with dual degrees in Linguistics and Spanish Language, gaining substantive knowledge of Arabic language and cross-lingual speaker recognition techniques, the topic of her undergraduate thesis. Since matriculating from graduate school, Ms. Richey has worked closely with the Intelligence and Defense communities in Washington, D.C. in the domain of advanced technology and applications of analytic methods to National Security mission requirements. She speaks regularly on advanced technology, digital transformation, and applications of Artificial Intelligence and Machine Learning (AI/ML) across a broad spectrum of problem domains. Ms. Richey also maintains her own consultancy, Merigold Analytics, and is the proud author of a local Virginia farm-to-table food blog, OneandahalfSlices.