

2018

A NETWORK BASED ANALYSIS OF THE INTERNATIONAL REFUGEE CRISIS USING GERGMS

Katherine E. Abramski
University of Rhode Island, keabra10@my.uri.edu

Follow this and additional works at: <https://digitalcommons.uri.edu/theses>

Terms of Use

All rights reserved under copyright.

Recommended Citation

Abramski, Katherine E., "A NETWORK BASED ANALYSIS OF THE INTERNATIONAL REFUGEE CRISIS USING GERGMS" (2018). *Open Access Master's Theses*. Paper 1293.
<https://digitalcommons.uri.edu/theses/1293>

This Thesis is brought to you by the University of Rhode Island. It has been accepted for inclusion in Open Access Master's Theses by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu. For permission to reuse copyrighted content, contact the author directly.

A NETWORK BASED ANALYSIS OF THE
INTERNATIONAL REFUGEE CRISIS USING GERGMS

BY

KATHERINE E. ABRAMSKI

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

STATISTICS

UNIVERSITY OF RHODE ISLAND

2018

MASTER OF SCIENCE THESIS
OF
KATHERINE E. ABRAMSKI

APPROVED:

Thesis Committee:

Major Professor Natallia Katenka

Marc Hutchison

Jing Wu

Nasser H. Zawia
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND
2018

ABSTRACT

The world is currently facing the worst migration crisis on record. Violent conflicts around the globe have forced over 65 million people to flee their homes, and receiving countries are struggling to support the massive influx of refugees. Lack of preparation and disorganization have only worsened the situation, and there is a pressing need to better understand refugee migration patterns in order to inform policy decisions and improve humanitarian efforts. In previous migration research, gravity models have been one of the classical methods for investigating determinants of migration, however this approach fails to take into account the interdependent nature of migration. To address this weakness, we apply statistical network analysis, which takes into account this interdependency, in order to quantify the influence of certain economic, political, social, and geographical factors on refugee migration. We create four different networks in order to investigate forced migration patterns surrounding four countries that are currently experiencing violent conflicts: Syria, Ukraine, the Democratic Republic of Congo (DRC), and Myanmar. Each network includes 12 nodes: the respective country of interest and the eleven countries hosting the most refugees from that country in 2015. Weighted directed edges represent the number of refugees from the origin country living in the host country in 2015. In order to quantify the influence of chosen factors on refugee migration in the context of the specific countries and conflicts of interest, we apply two different network models to each of

the four networks —the exponential random graph model (ERGM) adapted for binary edges, and the generalized exponential random graph model (GERGM) adapted for weighted edges. Our results indicate that the ERGM is a poor choice for modeling this specific problem since the thresholding required to coerce weighted edges into binary edges results in a failure to capture the vastly different magnitudes of refugee migration present in the networks. The GERGM proved to be a much better model. Our final GERGM produced vastly different results for each of the four networks, suggesting that refugee migration patterns differ greatly for different countries and conflicts. Our results also suggest that the influence of determinants of migration on refugee flow patterns differs greatly for out-migration and in-migration. We speculate that determinants of migration have a greater influence on out-migration than in-migration in the context of more recent conflicts. On the other hand, we speculate that determinants of migration have an influence on both out-migration and in-migration in the context of conflicts that have been ongoing for many years. This is likely a result of organized migration routes that have been established over many years.

ACKNOWLEDGMENTS

I would like to thank my advisors, Natallia Katenka, Marc Hutchison, and Jing Wu for providing guidance and support throughout my entire project. I would like to thank Ashley Buchanan for serving as my defense chair. I would also like to thank Gavino Puggioni, Katie Dio, Ayako Miura, and the rest of the Department of Computer Science and Statistics for providing continuous feedback and support throughout my two years at URI. Finally, I would like to thank my friends and family for helping me every step along the way.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGMENTS	iv
TABLE OF CONTENTS.....	v
LIST OF FIGURES	ix
LIST OF TABLES	ix
CHAPTER 1	1
INTRODUCTION	1
CHAPTER 2	5
BACKGROUND	5
<i>Previous research on migration.....</i>	<i>5</i>
<i>Network analysis</i>	<i>9</i>
<i>Data Description</i>	<i>12</i>
<i>Countries and conflicts</i>	<i>16</i>
<i>Justification for attributes</i>	<i>19</i>
CHAPTER 3	22
METHODS	22
<i>ERGM</i>	<i>22</i>
<i>GERGM</i>	<i>28</i>
CHAPTER 4	31
RESULTS	31
<i>Descriptive summary.....</i>	<i>31</i>
<i>ERGM.....</i>	<i>36</i>

<i>GERGM</i>	44
CHAPTER 5	52
CONCLUSION.....	52
<i>Implications of the results</i>	52
<i>Limitations</i>	55
<i>Future research</i>	59
BIBLIOGRAPHY	61

LIST OF FIGURES

Figure 1. A toy example of how the networks were built from the data, including the node and edge attributes.	15
Figure 2. The four networks of interest. Clockwise from top left: Syria, DRC, Myanmar, Ukraine. The country of interest is circled.	17
Figure 3. Diagram showing how different thresholds yield different network densities.	27
Figure 4. Histograms displaying the distribution of the edge attribute geodistance between countries for each of the four networks.	31
Figure 5. Boxplots showing the distribution of the different node attributes for the nodes included in each of the four networks.	32
Figure 6. Histograms displaying the distribution of edge weights (log refugees) for each of the four networks.	33
Figure 7. Bar graphs of density, transitivity, and reciprocity for the four different networks.	35
Figure 8. Histograms displaying the in-degree and out-degree distributions for each of the four networks.	36
Figure 9. The different network densities yielded by different threshold values.	38
Figure 10. Goodness-of-fit diagnostics plots for the ERGMs fit to the four networks. Clockwise from the top left: Syria, Ukraine, Myanmar, and DRC.	43
Figure 11. Plot of parameter estimates and credible intervals yielded by the	

GERGM fit to the Syria network.	46
Figure 12. Plot of parameter estimates and credible intervals yielded by the GERGM fit to the Ukraine network.	46
Figure 13. Plot of parameter estimates and credible intervals yielded by the GERGM fit to the DRC network.	47
Figure 14. Plot of parameter estimates and credible intervals yielded by the GERGM fit to the Myanmar network.	47
Figure 15. Diagnostics plots yielded by the GERGMs fit to each of the four networks. Clockwise from top left: Syria, Ukraine, Myanmar, DRC.	51

LIST OF TABLES

Table 1. A description of the node and edge attributes included in the study.	13
Table 2. The country codes corresponding to the networks of interest.	18
Table 3. A list of descriptive measures and their descriptions.....	34
Table 4. The number of refugees that correspond to each threshold cutoff.	39
Table 5. Parameter estimates and standard errors yielded by the ERGMs fit to the four networks.	40
Table 6. Summary of significance of parameter estimates yielded by the ERGMs fit to the four networks.....	42
Table 7. Parameter estimates and standard errors yielded by the GERGMs fit to the four networks.....	45
Table 8. Summary of significance of parameter estimates yielded by the GERGMs fit to the four networks.....	48

CHAPTER 1

INTRODUCTION

The world is currently facing the worst migration crisis on record. Violent wars, political conflicts, and economic crises across the globe have forced over 65.6 million people to flee their homes in an effort to escape the dire and often life-threatening conditions they face. Among the displaced, nearly 22.5 million have sought refuge outside their borders (UNHCR, 2018). This mass migration has put an incredible burden on receiving countries as they struggle to manage the massive influx of refugees. Neighboring countries have been hit the hardest, but further removed countries have been greatly impacted as well.

The negative consequences of the refugee crisis are widespread and severe, and in some cases, have infiltrated almost all aspects of everyday life. Some economic consequences faced by host countries include inflation, lower wages, increased competition for labor, and compromised public institutions such as education and health care (Rother et al., 2016; Masri and Srour, 2014). On the social spectrum, growing frustration and cultural differences have contributed to discrimination towards refugees and decreased social cohesion between refugees and their host communities (Rother et al., 2016; Masri and Srour, 2014). In Europe especially, the massive influx of refugees has sparked an intense anti-migrant movement

that has had a major political impact, fueling a shift towards right wing conservatism, as evidenced by recent political events such as Brexit (Podobnik et al., 2017). Tensions are high, and Europe is becoming increasingly divided on the migration issue as countries struggle to reach a consensus about how to distribute refugees (Collett, 2015). While receiving countries grapple with these political, economic, and social issues, millions of refugees are passing years of their life waiting in refugee camps, facing discrimination and low quality of life (Masri and Srour, 2014). In short, the situation is serious, and a solution seems very far off.

Current humanitarian efforts focus on short-term solutions to immediate problems. The European Commission's humanitarian and civil protection response provides aid for refugees and their host communities when their own emergency response capacities are overwhelmed, but it has fallen short (Collett, 2015). The European Commission has expressed the need for a kind of early warning system so that countries can prepare for a massive influx of migrants before it happens. According to the Migration Policy Institute, EU policymakers long expected an exponential increase in refugee arrivals, but little was done to prepare effectively. Furthermore they emphasize that policymakers need to narrow the gap between anticipating a foreign policy crisis and implementing policies that will prevent the type of chaos, disorganization, and turmoil that is now being experienced in Europe and around the globe (Collett, 2015).

This is the motivation that drives our research. The need to understand refugee flow patterns is now more important than ever. Being able to predict refugee migration flows could help mitigate the negative consequences of a massive influx of refugees, so in this study, we aim to lay the groundwork for future predictive models by investigating statistical network analysis as a tool for modeling forced migration.

In previous migration research, modified gravity models have been one of the classical methods for investigating determinants of migration. A major weakness of this approach is that it fails to take into account the interdependent nature of migration. To address this weakness, we apply statistical network analysis, which takes into account this interdependency, in order to quantify the influence of certain economic, political, social, and geographical factors on refugee migration. We create four different networks in order to investigate forced migration patterns surrounding four countries that are currently experiencing violent conflicts: Syria, Ukraine, the Democratic Republic of Congo (DRC), and Myanmar. Each network includes 12 nodes: the respective country of interest and the eleven countries hosting the most refugees from that country in 2015. Weighted directed edges represent the number of refugees from the origin country living in the host country in 2015. In order to quantify the influence of chosen factors on refugee migration in the context of the specific countries and conflicts of interest, we apply two different network models to each of the four networks—the exponential random graph model (ERGM) adapted

for binary edges, and an extension of the ERGM, the generalized exponential random graph model (GERGM) adapted for weighted edges. We explore the implications of our findings in the context of the countries and conflicts of interest, and we discuss the strengths and limitations of using both the ERGM and the GERGM for investigating this problem.

In the following section, we provide some background on previous research methods used in migration research. Mainly, we discuss the strengths and weaknesses of the modified gravity model, a commonly used method in migration research, and we discuss some research challenges in the field.

CHAPTER 2

BACKGROUND

Previous research on migration

Migration research addresses a broad range of questions. Who migrates? How does migration change over time? Where do migrants come from and where do they go? Why do people migrate? (Greenwood, 2005). In this study, we aim to investigate determinants of out-migration and in-migration in the context of specific countries and conflicts. We ask: Why do migrants come from where they come from, and why do they go where they go?

Naturally, factors that influence migration can be broadly grouped into two categories. Factors associated with the origin country or region that are related to the decision to migrate, such as extreme poverty or violence, are called push factors. Factors associated with the destination country or region that influence where migrants go, such as political stability or economic prosperity, are called pull factors (Langley et al., 2016). Additionally, intervening obstacles and personal factors such as migration policy and individual preferences can have a high degree of influence on migration (Lee, 1966).

One of the greatest challenges in migration research is accounting for all these different factors. A large body of research has been dedicated to

understanding determinants of migration, however many theories have focused on just one group of factors, such as economic factors or political factors, rather than considering the collective influence and interplay of many different factors together (de Haas, 2011). While some origin-specific and destination-specific factors may be relatively universal for the majority of migrants (i.e. migrants tend to migrate away from danger and towards safety), other factors such as intervening obstacles and personal factors can differ greatly from migrant to migrant. These factors are often difficult to measure, and their relative influence on migration is not well understood (de Haas, 2011).

One class of models that has commonly been used to assess the influence of many different factors simultaneously is the modified gravity model. This model is based on the original gravity model, which assumes that migration flows between two countries are proportional to their size (population or GDP) and inversely proportional to the geographical distance between them (Ramos, 2016). The equation for the gravity model is shown below:

$$M_{ij} = C \times \frac{P_i^{\beta_1} P_j^{\beta_2}}{D_{ij}^{\beta_3}}$$

M_{ij} is the migration between two counties, C is a constant, P_i and P_j reflect the sizes of the two countries respectively (population or GDP), and D_{ij} is the distance between them. β_1 , β_2 , and β_3 are the parameters to be estimated (Greenwood, 2005). The modified gravity model, which first

became popular in the 1960's, is an adaptation to this model that allows for the inclusion of other factors that are thought to influence migration. The equation is shown below:

$$\ln M_{ij} = \beta_0 + \beta_1 \ln P_i + \beta_2 \ln P_j + \beta_3 \ln D_{ij} + \sum \beta_k \ln X_k + \varepsilon_{ij}$$

Once again, M_{ij} is the migration between two countries, P_i and P_j reflect the sizes of the two countries respectively (population or GDP), and D_{ij} is the distance between them. The X_k term includes additional factors such as other economic, political, and social variables (Greenwood, 2005). The betas are the parameters to be estimated. The model can be easily estimated by ordinary least squares after a transformation into logarithmic form (Poot et al., 2016). For this reason, it has been very popular, not only in the past, but also in recent research.

In 2008, the Research Institute of Applied Economics used a gravity model for nearly 200 countries to analyze past and future trends in migration between the EU and EU neighboring countries between 1960 and 2010 (Ramos and Suriñach, 2013). They included various economic, political, social, and geographic variables in the analysis. In 2010, Karemera, Iwuagwu Oguledo, and Davis explored the influence of political, economic, and demographic factors on the size and composition of migration flows to Canada and the USA. They applied a modified gravity model that was specified and adjusted to include immigration regulations and characteristics specific to the origin and destination countries (Karemera et al., 2010). In 2017, Dedeoğlu and Deniz Genç fit a gravity

model to investigate determinants of emigration from Turkey to 31 European countries from 1960 to 2013 (Dedeoğlu and Deniz Genç, 2017).

While the modified gravity model is advantageous because it is easy to estimate and it allows the researcher to investigate many different factors simultaneously, it has some very limiting weaknesses. Since it is a regression model, it assumes independence between observations. This is a very strong assumption, one that is violated in the context of international migration since research has shown that migration flows between pairs of countries influence migration flows between other pairs of countries (Görlach and Motz, 2017). Failing to take into account this dependency factor could result in parameter estimates that inaccurately reflect the relationships between the dependent and independent variables (Cranmer and Desmarais, 2011).

For this reason, statistical network analysis is a natural way to represent refugee migration. It captures the interconnectedness of countries and allows the researcher to measure the influence of various factors on migration without assuming independence.

In the next section, we provide an overview of statistical network analysis, we introduce the family of exponential random graph models (ERGMs), and we review some previous applications of ERGM family models to questions related to migration.

Network analysis

Networks are complex systems that represent the relational structure of data. A network graph consists of a set of nodes and a set of edges, which may be directed or undirected, binary or valued. Subsequently, the elements of a network graph are dyadic pairs of nodes, which may or may not be connected by an edge (Kolaczyk, 2009). International migration can naturally be represented as a network where nodes represent countries and edges represent migration from one country to another. Additionally, networks allow for the inclusion of node and edge attributes, or covariates. These can be qualitative or quantitative characteristics that describe both the actors (nodes) and the relationships between them (edges). In the case of the present study, the economic, political, social, and geographical factors we are interested in investigating are included in the migration network as covariates. There are also several descriptive measures that are unique to networks. These measures, which are a direct result of the underlying structure of the network, provide interesting insight about the relational structure of the data. In the case of the present study, the descriptive measures of the network reflect how interdependencies among countries influence refugee migration. The covariates and descriptive measures of the network are the two main components used for statistical inference about networks, which we discuss later (Desmarais and Cranmer, 2017).

Perhaps the most important feature of network data is that relational ties between given pairs of actors depend on one or more of the other ties in

the network (Cranmer and Desmarais, 2011). As we discussed, this dependency is the reason that more traditional methods of statistical inference such as regression, which assume that the data are independent, are not suitable for modeling network data (Cranmer and Desmarais, 2011). This “dependency” factor is what makes network analysis an optimal choice for investigating questions related to migration between countries. Neglecting to consider interdependency can result in incorrect estimates of the covariates.

There are several well-established methods for statistical inference on networks. Some of the most commonly applied methods are exponential random graph models (ERGMs) and their extensions, latent space models, the quadratic assignment procedure, and stochastic actor-oriented models (Leifeld and Cranmer, 2014). After a careful review of the strengths and weaknesses of these models, we concluded that the ERGM family of models is the most suitable for this study. We apply the classical ERGM as well as the generalized ERGM (GERGM), which is an extension of the ERGM to networks with weighted edges. A detailed explanation of these models is provided in the Methods section.

The decision to use the ERGM family of models for this study was partially motivated by two recent studies, which both applied ERGM family models to model migration. In 2012, Desmarais and Cranmer applied a GERGM to a network of interstate migration in the USA from 2006 to 2007. The edges represented the difference in interstate migration from

2006 to 2007. Some of the covariates they included were unemployment, temperature, distance, income, and population. They included multiple structural components as well. For the covariates, temperature, population, and unemployment were found to be significant. For the structural components, they found clustering effects in the network and they found a lack of reciprocity (Desmarais and Cranmer, 2012).

In 2017, Windzio applied a longitudinal ERGM called a TERGM to a global migration network over four years. This model was used to understand and quantify the influence of geographic, demographic, economic, religious, linguistic, and historical factors on international migration. They found significant effects for geography, population, language, and religion. They also observed strong network structural effects that indicate a hierarchy in attractiveness for unobserved reasons (Windzio, 2017).

These studies serve as excellent resources for our research because they help inform model specification and interpretation, however our research is unique in that we investigate the migration of refugees specifically, which is distinct from voluntary displacement. Our research is also unique because we investigate factors that influence migration in the context of different conflicts and different groups of countries.

A more detailed description of the present study is provided in the following sections. This includes a description of the data that were used to build the networks, a description of the specific conflicts and countries that

we chose to investigate, and some justification for the different economic, political, social, and geographical factors that we chose to investigate in this study.

Data Description

A network graph $G = (V, E)$ is defined as a mathematical structure consisting of a set V of vertices (also commonly called nodes) and a set E of edges (also commonly called links), where elements of E are unordered pairs $\{u, v\}$ of distinct vertices $u, v \in V$ (Kolaczyk, 2009). We create four separate networks for the year 2015 where nodes represent countries around the world, and edges represent migration. Specifically, an edge from node A to node B represents the number of refugees originally from country A that reside in country B in 2015, therefore the edges are both directed and weighted.

The edge data, which represents refugees from one country living in another, are dyadic. These data were obtained from the UNHCR population database, which is open and available to the public. As defined by the UNHCR: “Refugees include individuals recognised under the 1951 Convention relating to the Status of Refugees; its 1967 Protocol; the 1969 OAU Convention Governing the Specific Aspects of Refugee Problems in Africa; those recognised in accordance with the UNHCR Statute; individuals granted complementary forms of protection; or those enjoying temporary protection. Since 2007, the refugee population also includes

people in a refugee-like situation.” (UNHCR, 2018). Dyadic data for all available countries were obtained. There were very few missing values, but those values that were missing were replaced with zeros under the assumption that very high levels of migration would not go unrecorded. Nonetheless, these data are likely not missing at random, since underdeveloped countries are more likely to keep poor records.

Variable name	Description	Source	Attribute
Log Population	Log of total population	World Bank	Node
Log GDP	Log GDP per capita (constant 2010 USD)	World Bank	Node
Unemployment Rate	Percentage of total workforce	World Bank	Node
Excluded Population	Percentage of ethnic minority population excluded from government	EPR	Node
Ethnic Fractionalization	Measure of ethnic diversity. The probability that two randomly chosen individuals will be of a different ethnicity	EPR	Node
Political Terror Scale	Level of political violence and terror that a country experiences (1-low, 5-high)	Political Terror Scale Project	Node
Geodistance	Geographical distance between countries (km/1000)	Correlates of War	Edge

Table 1. A description of the node and edge attributes included in the study.

Additionally, a number of node attributes and edge attributes were included in the networks. The data for the attributes were obtained from various sources. All data are for the year 2015. The variables, descriptions, and sources for the node and edge attributes can be found in Table 1. The

node attributes log population, log GDP, unemployment rate, excluded population, and ethnic fractionalization are numerical variables. Political terror scale is a categorical variable that has been scaled so that it is treated as a numerical variable. The edge attribute geodistance is a numerical variable as well and corresponds to the distance between pairs of countries.

Countries with missing geodistance data were not considered in the analysis, however the node attribute data had many missing values, and some countries had many missing values for many variables. We did not consider countries that had more than 30% missing values for the variables we considered, except in some special cases, in which we decided the countries were too important to exclude from the analysis (i.e. Syria). Missing values were then imputed using AMELIA, a program which performs multiple imputation on the dataset. AMELIA creates 10 different imputed datasets based on the available data and then takes the average of these 10 datasets to get a final imputed dataset. A limitation of AMELIA is that it has two assumptions. First it assumes that the data are multivariate normal, which in the case of the present study, they are not. Second it assumes that the data are missing at random. With refugee data, underdeveloped or war-torn countries are more likely to keep poor records. Therefore, these data are not missing at random, violating the assumption. The use of AMELIA despite the violation of these assumptions is a major limitation of our analysis.

From these data, we built the relevant networks. The figure below shows a toy example of how all data components are included in the networks that were built.

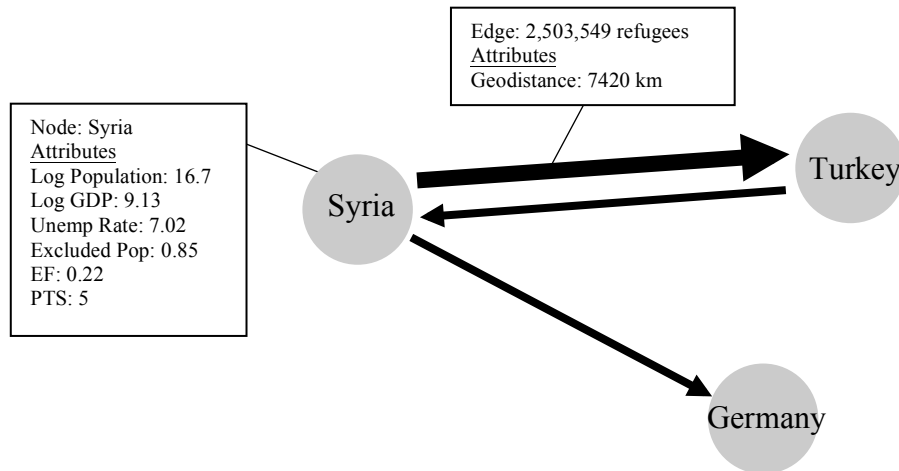


Figure 1. A toy example of how the networks were built from the data, including the node and edge attributes.

Each node represents a country and has node attributes, characteristics about that country. Directed edges represent the number of refugees living in one country from the origin country in 2015. Edges can go in both directions, and an edge can also be absent. The edges also have one attribute, which represents the geodistance between countries. When the four networks were built, since the distributions of the edges (refugees) were incredibly skewed right, we decided to take the log of the refugees to create the edges when building the networks in order to try to reduce some of the skewedness.

In the following section we discuss the countries and conflicts around which the four networks were built.

Countries and conflicts

The four networks were built around the following countries: Syria, Ukraine, the Democratic Republic of Congo (DRC), and Myanmar. We chose these countries because they are all currently experiencing different conflicts. The conflicts are of varying type, intensity, and duration, and are located in different parts of the world.

Syria: Syria is currently facing a violent civil war between the Syrian government—backed by Russia and Iran—and antigovernment rebel groups—backed by the United States, Saudi Arabia, Turkey, and others in the region. The conflict began in 2011 with protests against President Assad’s regime and since then over 400,000 people have been killed (Global Conflict Tracker, 2018). In 2015, there were 4.8 million Syrian refugees worldwide.

Ukraine: The crisis in Ukraine is a territorial dispute between Russian-backed separatists and the pro-EU Ukrainian military. It began in 2014 when Russian troops took control of the Crimean region (Global Conflict Tracker, 2018). In 2015, there were 321,418 refugees worldwide.

DRC: The current violence in DRC has its origins in the Second Congo War (1998-2003) during which government forces supported by Angola, Namibia, and Zimbabwe fought rebels backed by Uganda and Rwanda. Despite a peace deal and the formation of a transitional government in 2003, ongoing violence by armed groups against civilians has ensued as a result of weak governance and institutions as well as corruption (Global

Conflict Tracker, 2018). In 2015, there were 525,580 refugees from DRC worldwide.

Myanmar: The crisis in Myanmar is a sectarian dispute perpetrated by Buddhist nationalist groups targeting the Rhoingya, a highly persecuted Muslim minority group. In 2012, violence intensified and in 2015, there were 451,801 refugees from Myanmar worldwide (Global Conflict Tracker, 2018).

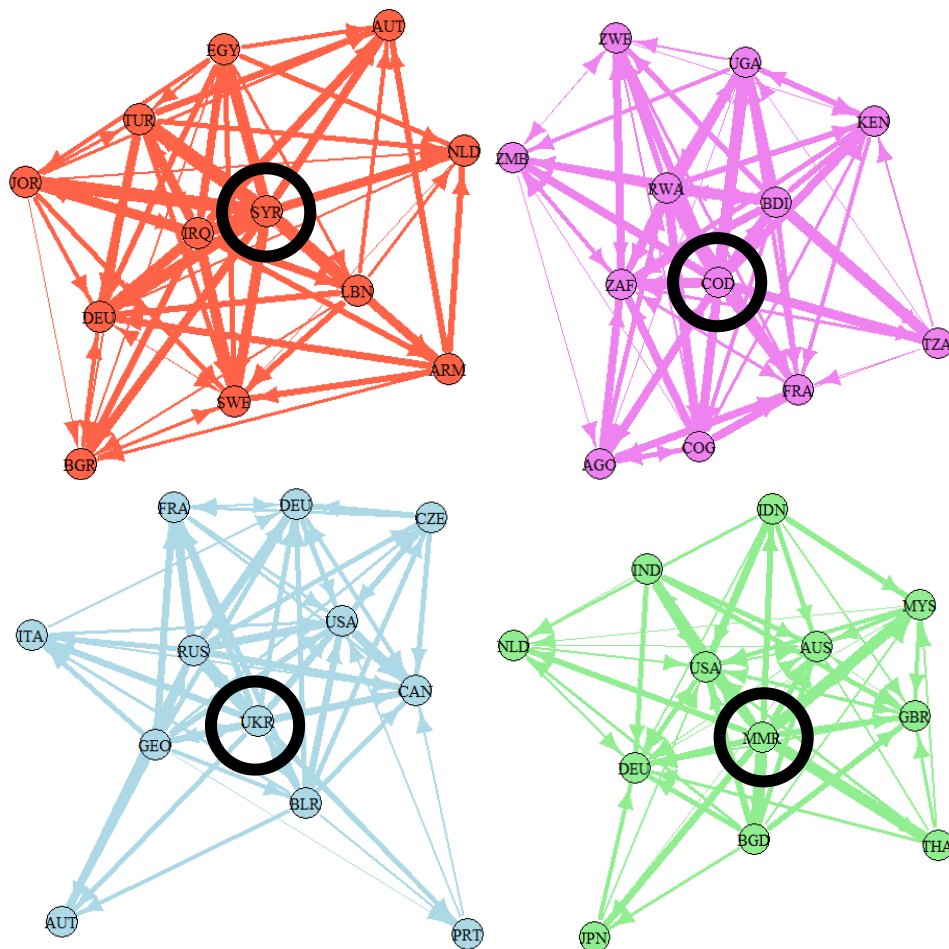


Figure 2. The four networks of interest. Clockwise from top left: Syria, DRC, Myanmar, Ukraine. The country of interest is circled.

The four networks, shown in Figure 2, were built around these four countries. Each network includes 12 nodes: the respective country of interest and the eleven countries hosting the most refugees from that country. The corresponding country codes are shown in Table 2.

<p>SYR – Syria ARM – Armenia AUT – Austria BGR – Bulgaria DEU - Germany EGY – Egypt IRQ - Iraq JOR – Jordan LBN – Lebanon NLD – The Netherlands SWE – Sweden TUR - Turkey</p>	<p>COD – Dem. Rep. of Congo AGO - Angola BDI - Burundi COG – Congo FRA – France KEN – Kenya RWA – Rwanda UGA – Uganda TZA - Tanzania ZAF – South Africa ZMB – Zambia ZWE – Zimbabwe</p>
<p>UKR - Ukraine AUT – Austria BLR – Belarus CAN - Canada CZE – Czech Republic DEU - Germany FRA - France GEO - Georgia ITA - Italy PRT - Portugal RUS – Russia USA – United States</p>	<p>MMR - Myanmar AUS - Australia BGD - Bangladesh DEU - Germany GBR – United Kingdom IND - India IDN – Indonesia JPN – Japan MYS - Malaysia NLD – The Netherlands THA – Thailand USA – United States</p>

Table 2. The country codes corresponding to the networks of interest.

We also considered creating the networks including the country of interest and its nearest neighbors, but we decided against this method since this resulted in the creation of sparser networks. A more detailed discussion about how we decided to build the networks is included in the discussion section.

Justification for attributes

The selection of factors to investigate in this analysis was motivated by the modified gravity model, previous research on determinants of migration, and limitations to the availability of data. Our decision to include log population in our model followed directly from the assumption of the modified gravity model that population is directly related to migration flows. Additionally, many studies have shown that economic factors are one of the main drivers of migration. Classical economic theories state that people make the choice to migrate when the benefit of moving generates the highest financial return on labor (Lewis, 1954). Before making the decision to migrate, a personal cost-benefit analysis is made to determine if a move would be beneficial or not (Langley et al., 2016). In short, economic migrants flee poverty to seek better employment opportunities in destinations that offer more economic stability. For this reason, we included factors log GDP and unemployment rate as economic factors in our study.

While migration due to economic factors is certainly significant, research has shown that the most prominent and powerful driver of international displacement is violent conflict because a high threat greatly increases the cost of staying (Langley et al., 2016; Ibáñez and Vélez, 2008). One study found that that the number of refugees leaving the origin country is proportional to the intensity of the threat (Moore and Shellman, 2004), and another study found that individuals fleeing greater threats to personal security are even more willing to use more dangerous migration routes

(Altai Consulting, 2015). For these migrants, the cost-benefit analysis has more to do with the cost of staying than the benefit of moving. For this reason, we included the political terror scale as a variable in our analysis. We also included excluded population and ethnic fractionalization as variables in the study since ethnic fragmentation has been found to be linked to increased levels of conflict (Akee et al., 2010).

Geographical proximity has also been shown to be positively related to migration between countries (Iqbal, 2007; Neumayer, 2005). The work of Day and White may provide some explanation for this. From a series of interviews with refugees in the UK, they found that migration is often a process with two stages. First, migrants tend to flee to a neighboring country to seek immediate safety, and then they seek a more permanent settlement (Day and White, 2001). We have included the geographical distance between countries in our model to investigate this factor.

It is important to note that there are several other factors that have been shown to influence refugee migration, some of which are difficult to measure. For example, community networks have been shown to be a strong pull factor since having connections in the destination country decreases the cost of migrating (Langley et al., 2016). Community networks, however, are difficult to measure quantitatively, and it also takes a considerable amount of time for these community networks to develop, so they may not emerge until several years after the outbreak of an armed conflict.

The influence of migration policy on migration is slightly less understood. Several studies have shown that restrictive migration policies are effective at reducing migration to certain countries, but in dire situations, policies are often relaxed since the threat migrants face becomes more significant than the policies in place (Schaeffer, 2010; Thielemann, 2006). This makes migration policy a challenging factor to understand, since it is difficult to measure the extent to which migration policies are relaxed, and for which populations. Migration policy is also a difficult variable to measure, thus, it has not been included in the present study.

It is important to note that in many cases, migrants fleeing serious threats to their safety are unable to choose their destination country, and where they go is often a result of chance rather than planning (Crawley, 2010). For the majority of these migrants, the main priority is to reach a safe place, and they are often completely unaware of the welfare situation in the destination country (Robinson and Segrott, 2002). In this way, migrant preferences are often rendered obsolete since where they end up may be largely determined by factors out of their control. This creates a challenge for researchers, since it cannot be known if a refugee is in a certain country by choice or because of outside factors.

In the following section, we provide a detailed description of the classical ERGM and the GERGM.

CHAPTER 3

METHODS

ERGM

In this section, we describe in detail the classical ERGM, which lays the foundation for the GERGM. We discuss how and why it is useful, its analytical form, specification and interpretation, and limitations. We then describe how the classical ERGM can be extended to the GERGM for networks with valued edges.

The development of the ERGM (Wasserman and Pattison, 1996) was a great advancement in network science since it provided a long-anticipated solution to the dependency problem that arises with classical statistical models (Cranmer and Desmarais, 2011). Similar to regression, the ERGM can account for the effects of covariates on the status of relationships between actors. In other words, the influence of the node and edge attributes on the structure of the network can be measured. Additionally, the ERGM can model the prominence and significance of structural dependencies of the network, such as the descriptive measures previously mentioned (Desmarais and Cranmer, 2017). In this way, the researcher can test specific hypotheses about how certain network structures drive the formation of the network. Even if the researcher is uncertain about the types of interdependencies that underlie the formation of the network,

the model can shed light on specific relational patterns inherent in the data that may be intuitively difficult to recognize (Desmarais and Cranmer, 2017).

In the context of the ERGM, the observed network can be thought of as a single observation from a multivariate distribution where many other realizations of the network are possible (Cranmer and Desmarais, 2011). Then the goal is to select features of the observed network that differentiate it from a random draw from the uniform distribution of all other possible networks with the same number of nodes that could be observed (Cranmer, Desmarais, and Menninga, 2012). These features are included in a set of statistics computed on the network. Then the parameters of the model are estimated to maximize the likelihood of observing the network of interest, Y . These parameters tell the researcher how the covariates as well as the inherent network structures drive the formation of the network (Cranmer and Desmarais, 2011).

If Y^* is a random network of n nodes, then the probability of observing the network Y rather than all other possible networks can be expressed as a function of the set of statistics:

$$P_{\theta}(Y^* = Y) = \frac{\exp\{\theta^T \Gamma(Y)\}}{\sum_{\text{all graphs } y^*} \exp\{\theta^T \Gamma(Y^*)\}}$$

where θ^T is a vector of parameters and $\Gamma(Y)$ is a vector of networks statistics (Cranmer, Desmarais, and Menninga, 2012). The statistics contain information about the covariates and the network structures, as specified by the researcher.

Node and edge covariates, X_n and X_d respectively, are included in the vector of statistics as $\Gamma_{X_n}(Y, X_n) = \sum_{i \neq j} X_i X_j Y_{ij}$ and $\Gamma_{X_d}(Y, X_d) = \sum_{i \neq j} X_{ij} Y_{ij}$ (Cranmer and Desmarais, 2011). The covariates should be chosen such that high values of that covariate will either decrease or increase the probability of observing an edge. In this way, the ERGM is similar to regression because it can quantify the effect of some covariate on an outcome. If some covariate X has a positive effect, then a higher value of X should increase the probability of an edge in Y .

The network structures are included in the vector of statistics similarly. For example, reciprocity is accommodated as $\Gamma_R(Y) = \sum_{i < j} Y_{ij} Y_{ji}$ while in-two-stars and out-two-stars are accommodated as $\Gamma_{in}(Y) = \sum_i \sum_{j < k \neq i} X_{ji} X_{ki}$ and $\Gamma_{out}(Y) = \sum_i \sum_{j < k \neq i} X_{ij} X_{ik}$ (Denny, 2016). The challenge lies in choosing the network structures to include in the vector of statistics. The researcher must choose the network structures that are believed to increase the probability of observing the observed network based on what drives its formation. For example, in the case of our migration networks, we understand that migrants tend to leave countries that are very conflicted, so we would not expect there to be a lot of migration into countries that are also experiencing a lot of out-migration. Therefore, we include reciprocity as a network structure in the set of statistics since we expect the observed network to have much less reciprocity than a random network drawn from a uniform distribution of networks with the same number of nodes.

The parameters of the network are estimated for each statistic included in the model. This is done by maximizing the likelihood function, however the exact computation of the likelihood function is too computationally demanding for any network with more than a few nodes because it requires the summation over all possible network configurations, which is $2^{\binom{N}{2}}$ for an undirected network with N nodes (Cranmer and Desmarais, 2011). Therefore, the likelihood function must be approximated since there is no closed form solution. The two most common methods of approximation are maximum pseudolikelihood (Frank and Strauss, 1986) and Markov chain Monte Carlo (MCMC) maximum likelihood (Geyer and Thompson, 1992).

The estimated parameter values can be interpreted similar to regression coefficients. If a parameter estimate is significantly different from zero, then we can conclude that its corresponding statistic significantly effects the probability of observing a particular instance of that network, controlling for the other statistics included in the model (Cranmer, Desmarais, and Menninga, 2012). In other words, we can conclude that the patterns observed in the network of interest did not occur by chance. Each parameter estimate can be interpreted as the log-odds increase (positive) or decrease (negative) in the probability of forming an edge. In this way, by looking at the parameter values of the model applied to the network of international refugee migration, we can quantify the influence of each

individual factor on the formation of the network as well as understand which interdependent structures drive the formation of the network.

The ERGM is a flexible tool because it only relies on two assumptions. First, it assumes that the network statistics calculated on the observed network are the expected values of those statistics across all possible graphs: $E(\Gamma_i) = \Gamma_i$. This is a strong assumption, but in many cases the observed network is the only network we can possibly observe, so it is the best estimate that we have (Cranmer and Desmarais, 2011). Second, we assume that the model is specified correctly, meaning that only network statistics chosen by the researcher influence the probability that Y is observed (Cranmer and Desmarais, 2011). Therein lies the challenge.

Some limitations of the ERGM should be noted. One problem that arises when the model fits the data poorly is degeneracy. This usually results from specifying a model that is so unlikely to have generated the network that the ERGM estimates cannot be computed (Cranmer and Desmarais, 2011). If the researcher runs into the problem of degeneracy, it is a sign that chosen components of the model must be reconsidered. Degeneracy is not so much a limitation, rather a red flag that the model has failed and needs to be specified differently. Another issue that the researcher may encounter is missing data. This may be a greater problem for ERGMs than it is for other more traditional models because effective multiple imputation may be more difficult to achieve (Cranmer and Desmarais, 2011). One way around this problem is using multiple

imputation on the edge-list representation of the network (Cranmer and Desmarais, 2011).

Another limitation of the traditional ERGM is that it is only suitable for networks with binary edges. One way around this limitation is thresholding the valued edges to coerce the valued-edge network into a binary network that can be modeled with an ERGM. Different threshold values yield different numbers of edges and thus networks of varying densities (Cranmer and Desmarais, 2011). This can lead to problems of degeneracy, but it can also lead to the loss of important information about the data. Figure 3 provides a good visual representation of how thresholding on different values can lead to different graph densities.

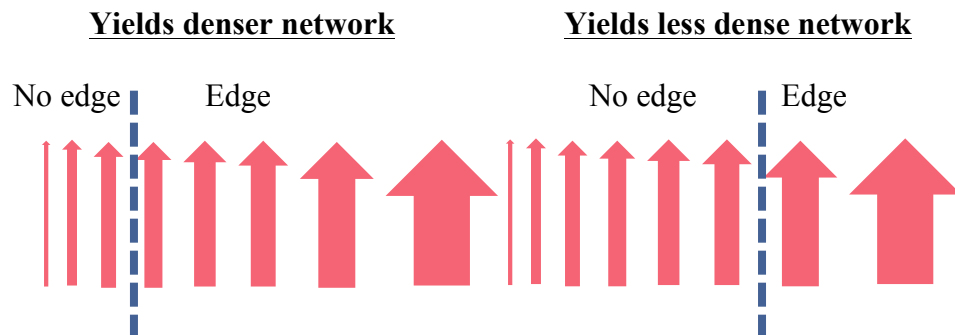


Figure 3. Diagram showing how different thresholds yield different network densities.

A lower threshold (shown on the left) results in more edges and subsequently a higher network density while a higher threshold (shown on the right) results in less edges and subsequently a lower network density. Choosing the right threshold for the model can prove a difficult task, especially when there is high variation among the edges.

In 2012 Desmarais and Cranmer developed the GERGM to address this problem so that valued-edge networks do not have to be coerced into binary networks, rather they can remain continuous. In the case of the migration network, there are varying degrees of migration so thresholding to create a binary network can lead to the loss of important information. For this reason we apply the GERGM as well as the classical ERGM.

GERGM

The idea behind the GERGM is the same as that of the ERGM, however the procedures for specification and estimation are slightly different. Specification is a two-step process. First, a joint distribution that captures the structure and interdependence of the observed network Y is defined on a restricted network configuration, $X \in [0,1]^m$ where m is the total number of directed edges between nodes (Desmarais and Cranmer, 2012; Wilson et al., 2016). Note that X has the same vertices as Y , but the edge values are continuous and bounded between zero and one. Then, X is transformed onto the support of Y through an appropriate transformation function, which creates a probability model for Y (Desmarais and Cranmer, 2012; Wilson et al., 2016).

In the first step, a set of network statistics, \mathbf{h} is defined to contain information about the covariates and the network structures, as with the ERGM. Then a probability distribution for X is defined by modifying the

ERGM formula to have a convergent sum in the denominator for a bounded network (Desmarais and Cranmer, 2012; Wilson et al., 2016):

$$f_X(X, \theta) = \frac{\exp(\theta^T \mathbf{h}(X))}{\int_{[0,1]^m} \exp(\theta^T \mathbf{h}(Z)) dZ}$$

where $\theta \in \mathbb{R}^p$ is the vector of parameters and $\mathbf{h}: [0,1]^m \rightarrow \mathbb{R}^p$ is formulated to represent the joint features of Y in the distribution of X (Desmarais and Cranmer, 2012; Wilson et al., 2016). This specification resembles that of the ERGM except the edges are now modeled as continuous taking values between zero and one (Desmarais and Cranmer, 2012; Wilson et al., 2016).

In the second step, the restricted network X is transformed onto the support of the observed network Y by applying a parameterized one-to-one monotone increasing transformation function $T^{-1}: [0,1]^m \rightarrow \mathbb{R}^m$ to the m edges of the restricted network. Specifically, for each pair of distinct nodes i, j we have $Y_{ij} = T_{ij}^{-1}(X, \beta)$ where $\beta \in \mathbb{R}^k$ parameterizes the transformation to capture the marginal features of Y (Wilson et al., 2016). This transformation allows for the specification of the GERGM such that the basic structure, strength, and flexibility of the classical ERGM are maintained, only now the vector of statistics \mathbf{h} is specified on a transformation of the network rather than the network in its observed form (Desmarais and Cranmer, 2012). The GERGM, which is the pdf of Y , can be written (Desmarais and Cranmer, 2012; Wilson et al., 2016):

$$f_Y(Y, \theta, \beta) = \frac{\exp(\theta^T \mathbf{h}(T(Y, \beta)))}{\int_{[0,1]^m} \exp(\theta^T \mathbf{h}(Z)) dZ} \prod_{ij} t_{ij}(Y, \beta)$$

where $t_{ij}(Y, \beta) = \frac{dT_{ij}(Y, \beta)}{dY_{ij}}$. When choosing the transformation T^{-1} , the distribution of the data should be considered, and while there is flexibility in this choice, it is wise to select a transformation such that T_{ij}^{-1} is an inverse cdf as it leads to beneficial properties (Wilson et al., 2016). As with the ERGM, the parameters of the GERGM are estimated by maximizing the likelihood function. Once again, the exact computation of the likelihood function is almost always too computationally demanding, so the likelihood must be approximated. This can be done with MCMC using Gibbs sampling (Desmarais and Cranmer, 2012) or alternatively Metropolis-Hastings methods can be used (Wilson et al., 2016). Interpretation of the parameter estimates is synonymous to interpretation with the classical ERGM.

As with the classical ERGM, the main challenge of the GERGM is the correct specification of the model, that is, the specification of the covariates and structural components to be included in the statistics vector, as well as the selection of the transformation function. In this study, the selection of these components was determined based on the properties of the data, previous research, and knowledge about the problem of interest.

In the next section, we provide descriptive summary statistics of the networks, we explain how the different models, both ERGM and GERGM were specified, and we present the results of both the ERGM and the GERGM.

CHAPTER 4

RESULTS

Descriptive summary

In this section, we provide descriptive summary statistics about the networks and covariates included in the networks. Figure 4 displays the distribution of the edge attribute, geodistance (in thousands of kilometers) between countries, for each of the four networks.

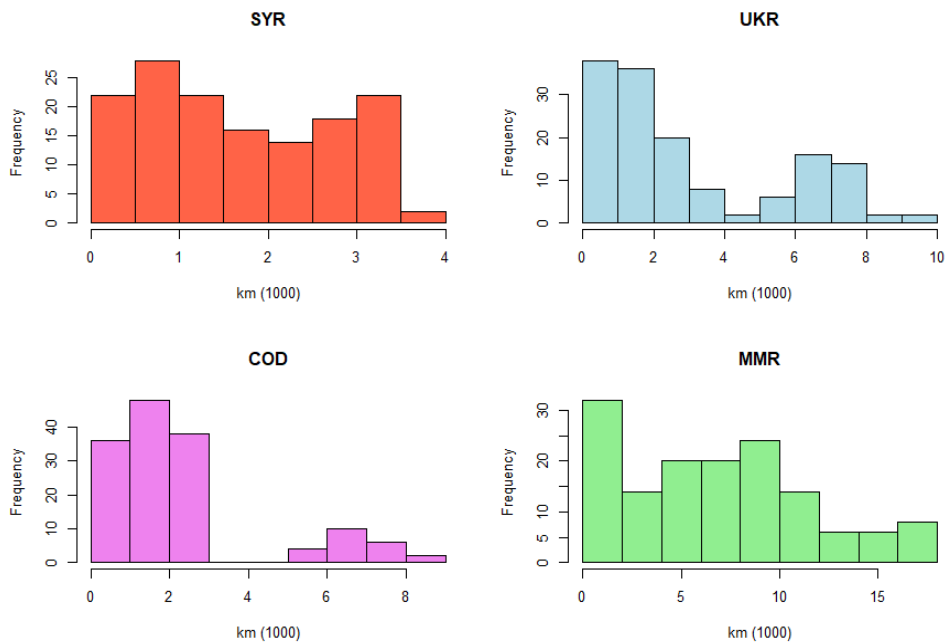


Figure 4. Histograms displaying the distribution of the edge attribute geodistance between countries for each of the four networks.

There is a clear bimodal pattern in the distributions of geodistance for Syria, Ukraine, and DRC. This suggests that the countries that comprise each of the networks can be divided into two groups: countries that are

relatively close to each other and close to the country of interest (reflected in the left side of the histogram) and countries that are further away from each other and further from the country of interest (reflected in the right side of the histogram). This suggests that the refugees from these three countries (Syria, Ukraine, and DRC) are split between neighboring countries and non-neighboring countries.

The boxplots in Figure 5 show the distribution of the different node attributes for the nodes included in each of the four networks. Each plot corresponds to a different node attribute – log population, log GDP, unemployment rate, excluded population, ethnic fractionalization, and political terror scale.

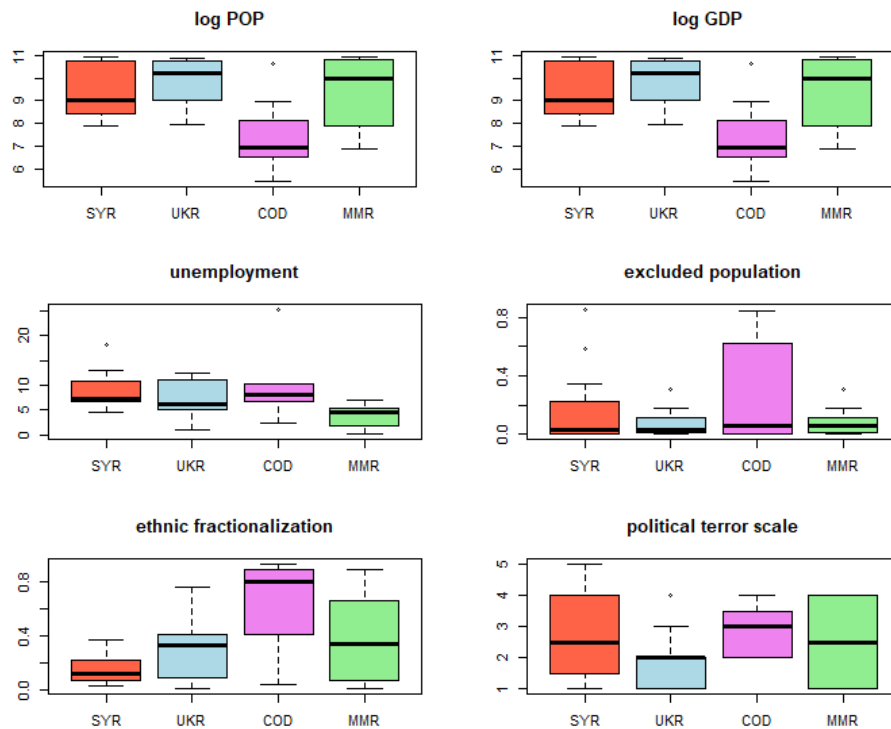


Figure 5. Boxplots showing the distribution of the different node attributes for the nodes included in each of the four networks.

From the boxplots, we can see that there is a lot of variation among the node attributes in the Myanmar network as well as a lot of variation in ethnic fractionalization and excluded population for DRC. This suggests that the countries included in the network are very different with respect to these specific attributes. On the other hand, there is little variation among the node attributes in the Ukraine network, suggesting that the countries included in the network are similar with respect to these specific attributes.

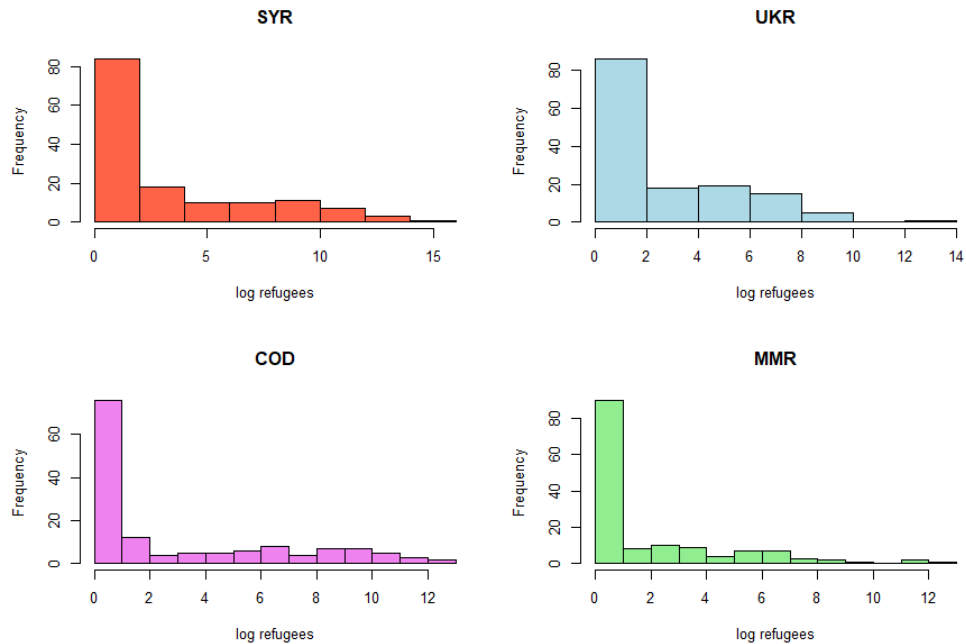


Figure 6. Histograms displaying the distribution of edge weights (log refugees) for each of the four networks.

The histograms in Figure 6 show the distribution of the edge weights for each network. Clearly, the distributions are very heavily skewed to the right. It is important to note that these are the log refugees, not the original number of refugees, for which the distributions are even more skewed to the right. It can also be noted that these distributions are zero inflated, reflecting

the absence of edges in the networks. Note, the absence of an edge means that there are no refugees from country A living in country B in 2015.

We now shift our focus to descriptive measures about the characteristics of the networks rather than the attributes. These measures provide interesting insight about the relational structure of the data. The five descriptive measures that we investigated are listed in the Table 3 (Kolaczyk, 2009).

Descriptive Measure	Description
Density	The proportion of realized edges to the total possible edges in a network
Transitivity	The proportion of closed triangles to connected triples in the network
Reciprocity	The likelihood that two nodes in a directed network are mutually connected
Degree (in and out)	The number of edges going into or out of a node

Table 3. A list of descriptive measures and their descriptions.

Figure 7 displays bar graphs that compare the density, transitivity, and reciprocity of the four networks. Looking at the plots, it appears that these descriptive measures are related. The network density is highest for the Syria and DRC networks, followed by the Ukraine network, and then the Myanmar network. All densities are around 0.5, suggesting that around 50% of possible ties have been realized. There is a similar pattern for transitivity, with values around 0.9. This high transitivity (many closed triangles) is not surprising. Consider a country, A, which is very conflicted, another country, B, which is moderately conflicted, and a country, C, which

is peaceful. Most likely, there will be refugees in country C from countries A and B but there may also be refugees in country B from country A since refugees tend to move to more peaceful countries. Finally, reciprocity is highest for DRC and lowest for Myanmar while it is around the same for Syria and Ukraine. All values are around 0.3, which is quite low, but it is surprising since we would expect a negative measure of reciprocity since we would not expect bidirectional migration between countries. These three measures reflect the “connectedness” of the networks.

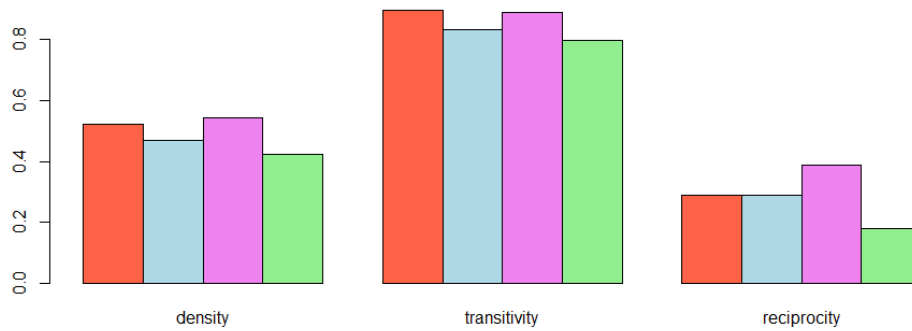


Figure 7. Bar graphs of density, transitivity, and reciprocity for the four different networks.

Figure 8 shows the in and out degree distributions for each network. There is a slight bi-modal pattern present in a few of the histograms which suggests that there are some dominant “sender” countries and some dominant “receiver” countries.

In the next section, we describe how the ERGM was fit and we present the results obtained from fitting the ERGM to each of the four networks.

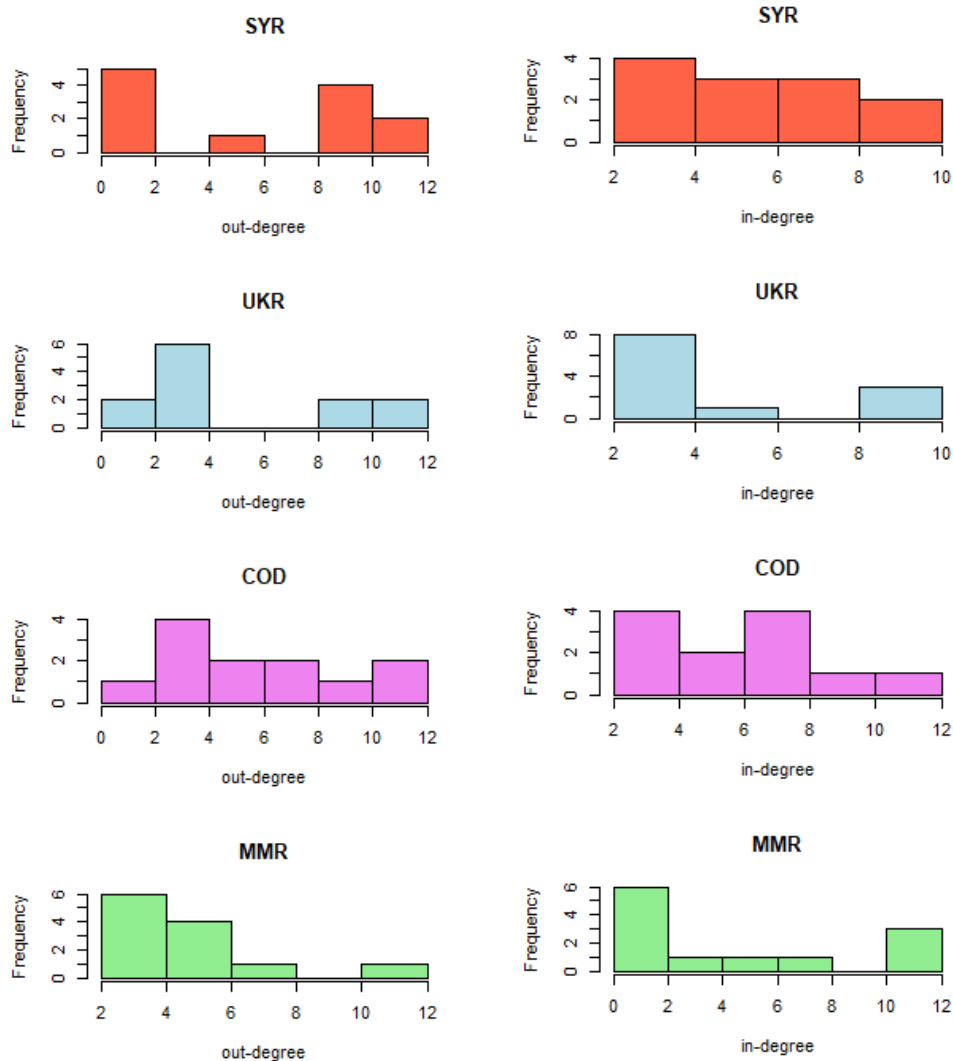


Figure 8. Histograms displaying the in-degree and out-degree distributions for each of the four networks.

ERGM

Model specification includes deciding which statistics to include in the model. We included three network statistics, six node covariates, and one edge covariate. The network statistics we included were mutual dyads also called reciprocity, which measures the extent to which there is mutual migration between a pair of countries, in-stars, which measures “popularity”, or the tendency of some countries to receive more refugees

than others, and out-stars, which measures “sociality”, or the tendency of some countries to send more refugees than others. The node covariates log population, log GDP, unemployment rate, ethnic fractionalization, excluded population, and political terror scale were included as both sender and receiver effects to measure their respective influence on out-migration and in-migration. Finally, geodistance was included as the only edge covariate, however distance (km) was divided by 1000 to reduce the scale. All edges – the number of refugees – were included as the log of the refugees in order to reduce the skewedness.

One major challenge in fitting the ERGM was deciding how to threshold the edges in order to coerce the valued-edge networks into binary networks. As discussed earlier, different threshold values yield different network densities, and thresholding can also lead to the loss of important data. The challenge, then, is to choose a threshold such that a desirable density is obtained while still preserving the meaning of the network data.

In the case of the migration networks investigated in this study, this was especially a problem since the data are so incredibly skewed. Considering the edges, the vast majority were zeros or otherwise very low numbers of refugees. On the other hand, there were a few very heavy edges, such as the edge from Syria to Turkey, as well as some moderately heavy edges, such as the edge from Syria to Germany. The problem with thresholding is that it forces these edges into two groups. For instance, if the threshold value is set too low, the great significance of the migration from

Syria to Turkey (over 2 million people) would be lost, since this edge would be valued the same as an edge from Syria to Germany. Instead, a threshold value that is set too high would completely fail to capture the significance of the migration from Syria to Germany, which is nonetheless very significant. Additionally, a threshold value that is too high would yield a network that is too sparse (not dense enough), to effectively run an ERGM.

In order to investigate different thresholds, we looked at the network densities that resulted from thresholding the networks at different threshold values, shown in Figure 9. Keep in mind that the threshold values are on the scale of log refugees, and range from 0 to 15. Table 4 shows the equivalent raw number of refugees that correspond to each threshold value.

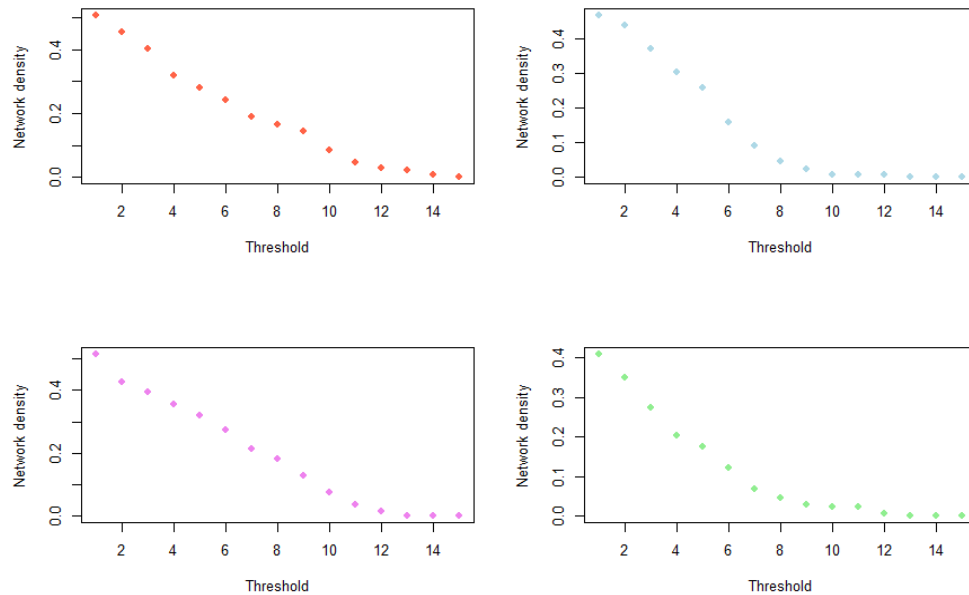


Figure 9. The different network densities yielded by different threshold values.

Log refugees	Refugees
2	7.38
4	54.6
6	403
8	2,981
10	22,026
12	162,754
14	1,202,604

Table 4. The number of refugees that correspond to each threshold cutoff.

From Figure 9, we can see that even at a threshold of 1, which would preserve all edges, the density of the networks are still around 50%. The density dramatically drops when the threshold is increased. The problem is that in order to preserve the significance of high levels of migration, a very high threshold is needed. A threshold of 10, which is a high threshold, considering it yields densities of less than 10% for all four networks, corresponds to just 22,026 refugees, which pales in comparison to the 2.5 million Syrian refugees, which are currently residing in Turkey. Using this threshold would completely eliminate this information from the network, while using a higher threshold would result in an even lower density.

When fitting the ERGM to the different networks, we explored different threshold values. All models were degenerate when fitted to networks that were created from a threshold of 6 or more. The best models based on the AIC were chosen as the final models. For Syria, a threshold of 3 (20 refugees) yielded the best model, for Ukraine, a threshold of 3, for DRC, a threshold of 5 (148 refugees), and for Myanmar, a threshold of 4

(54 refugees). The output in Table 5 shows the parameter estimates and standard errors produced by the models for each network.

	Syria		Ukraine		DRC		Myanmar	
Threshold level	Log 3 = 20 refugees		Log 3 – 20 refugees		Log 5 = 148 refugees		Log 4 = 54 refugees	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
intercept	-38.1	25.3	-37.9	25.1	-40.2	29.0	-25.6	15.0
Mutual	-0.14	1.58	0.26	1.46	3.75	2.03	---	---
Out-stars	0.21	0.25	-0.34	0.63	-0.11	0.45	0.34	0.19
In-stars	-3.52	1.90	0.02	0.55	-0.16	0.74	0.54	0.23
log population sender	-0.25	0.39	2.88	1.92	1.26	1.22	0.14	0.28
log GDP sender	-1.30	0.71	-6.46	3.83	-2.66	1.59	-0.53	0.34
unemployment sender	-0.02	0.14	-0.01	0.21	0.33	0.28	-0.20	0.22
excluded population sender	1.12	2.10	13.64	9.27	13.72	6.21	18.38	7.10
ethnic fractionalization sender	5.30	5.19	5.59	4.26	-5.97	4.10	0.82	1.33
political terror scale sender	1.25	0.81	-1.28	2.52	---	---	---	---
log population receiver	-0.14	0.92	1.44	0.62	1.27	0.85	0.44	0.43
log GDP receiver	7.56	3.32	2.71	1.35	1.78	0.63	1.33	0.64
unemployment receiver	-0.40	0.35	0.05	0.12	0.16	0.19	0.25	0.27
excluded population receiver	0.80	4.11	4.05	5.70	-2.43	2.40	0.54	6.75
ethnic fractionalization receiver	-27.9	17.9	8.29	3.23	-0.37	2.43	2.33	1.92
political terror scale receiver	0.35	1.46	-0.82	0.82	---	---	---	---
geodistance	-0.28	0.64	-0.61	0.27	-1.24	0.56	-0.10	0.11

Table 5. Parameter estimates and standard errors yielded by the ERGMs fit to the four networks.

Each parameter estimate can be interpreted as the change in the (log-odds) likelihood of a tie for a unit change in a predictor given the realized attributes and the rest of the network. Predictors are network-level statistics

that represent Markovian processes, so we can think about their changes locally. For example, in the above table, the estimate of -0.28 on geodistance for the Syria network means that the log-odds probability of forming an edge decreases by 0.28 for every unit of distance between the nodes/countries. In this way, positive coefficients reflect variables that influence the formation of the network positively (facilitate the formation of ties) while negative coefficients reflect variables that influence the formation of the network negatively (hinder the formation of ties). Consider the variable log GDP receiver, for example, which has a positive coefficient value for all four networks. This implies that if a node has high log GDP, it is more likely to have an edge going into it. Conversely, the variable ethnic fractionalization receiver for Syria has a negative coefficient, which implies that if a country has high ethnic fractionalization, it is less likely to have an edge going into it.

Whether the variables have significant effects or not depends on the magnitude of the estimates and the standard errors. In Table 6, the significance of the various parameter estimates are represented by different colors in the table. White rectangles represent non-significant coefficients, while red rectangles represent negative coefficient values and green rectangles represent positive coefficient values. The lightest, middle, and darkest shades represents coefficients that are significant at the 10%, 5%, and 1% levels respectively. From this table, we can see a pattern with the variables for log GDP. Log GDP sender has positive coefficients for the

Syria, Ukraine, and DRC networks while log GDP receiver has a negative coefficient for all networks. These results suggest that, within these networks, if a country has higher GDP, it is more likely to have more in-migration and if a country has lower GDP, it is more likely to have more out-migration. This is consistent with what we would expect, since a poor economic situation is a main driver of forced migration. The coefficients for geodistance for Ukraine and DRC are very significant and negative. This is also consistent with what we would expect since we would expect less migration between countries that are further away from each other. There do not appear to be any other clear patterns in the results of the parameter estimates.

	Syria	Ukraine	DRC	Myanmar
intercept				
Mutual				
Out-stars				
In-stars				
log population sender				
log GDP sender				
unemployment sender				
excluded population sender				
ethnic fractionalization sender				
political terror scale sender				
log population receiver				
log GDP receiver				
unemployment receiver				
excluded population receiver				
ethnic fractionalization receiver				
political terror scale receiver				
geodistance				

Table 6. Summary of significance of parameter estimates yielded by the ERGMs fit to the four networks.

The plots in Figure 10 show the goodness-of-fit diagnostics for the four ERGMs fitted to the four networks. The boxplots display the variation of the statistics calculated on the networks simulated from the parameter estimates, while the solid black line displays the network statistics calculated on the observed network. The network statistics calculated on the observed network should be approximately the expected values of those statistics across all possible graphs. As we can see, the plots reveal a relatively good fit, however, the results have little significance in the context of the problem since the thresholds that were used to coerce the weighted edges into binary edges were so low, resulting in the loss of almost all significant information about the networks. This is motivation for using the GERGM, the model that is adapted for weighted edges. With this model, the significant information about the networks is preserved and we can still estimate the influence of the same parameters.

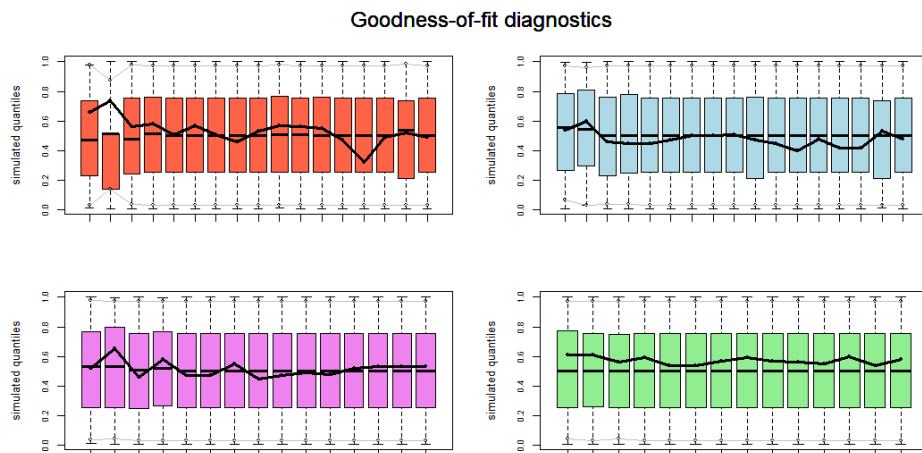


Figure 10. Goodness-of-fit diagnostics plots for the ERGMs fit to the four networks. Clockwise from the top left: Syria, Ukraine, Myanmar, and DRC.

GERGM

The GERGMs were fit including the same statistics as the ERGMs that we fitted. A main component of the GERGM, which is not present in the ERGM, is the monotone increasing transformation function. The Cauchy and the log Cauchy transformation functions are both suitable for skewed data, while the log Cauchy function is only suitable for non-zero data. Since the distributions of the edge weights in the networks are incredibly skewed and include zeros, the Cauchy transformation seemed suitable. However, we found that the log Cauchy transformation, which is only suitable for non-zero edges, worked better than the Cauchy transformation after adding 0.01 to all edges to make them non-zero, so this is what we did. The Gaussian transformation was another option, but this transformation is not suitable for highly skewed data.

The results of the GERGMs fitted to the Syria, Ukraine, and DRC networks are shown in Table 7. For each country, the parameter estimates and standard errors are shown in the chart.

The interpretation of the parameter estimates is similar to that of the ERGM. Significant parameter estimates affect the width of the edge conditional on the rest of the network. For instance, log GDP sender has a negative parameter estimate for all four networks, so a node with higher GDP is expected to have a narrower edge going out of it. In other words, it is expected to have less out-migration than a country with lower GDP, all else equal. The extent to which the edge is narrower or wider depends on

the magnitude of the coefficient and is a function of the transformation function.

	Syria		Ukraine		DRC		Myanmar	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE
intercept	-36.0	2.68	-5.98	3.54	-31.6	19.10	-18.8	2.40
Mutual	6.98	2.04	-3.27	1.74	-0.13	1.47	5.82	1.73
Out-stars	1.56	0.11	0.60	0.30	0.62	0.22	-2.00	0.22
In-stars	-3.00	0.31	0.02	0.31	-0.72	0.27	1.43	0.10
log population sender	0.69	0.05	0.00	0.06	0.41	0.32	0.02	0.04
log GDP sender	-8.13	0.16	-9.14	0.28	-3.16	0.69	-0.03	0.12
unemployment sender	1.37	0.07	-0.95	0.09	0.05	0.26	-0.15	0.05
excluded population sender	1.63	0.05	0.45	0.06	2.00	0.21	0.28	0.05
ethnic fractionalization sender	2.22	0.07	0.46	0.06	-1.66	0.21	-0.09	0.04
political terror scale sender	-1.63	0.13	0.12	0.09	1.83	0.27	---	---
log population receiver	0.38	0.10	0.67	0.29	1.81	1.36	0.96	0.40
log GDP receiver	-0.03	0.11	0.48	0.28	1.15	0.64	7.83	0.17
unemployment receiver	-0.05	0.04	0.06	0.06	1.38	0.36	0.32	0.11
excluded population receiver	0.04	0.03	0.09	0.06	0.53	0.22	0.90	0.08
ethnic fractionalization receiver	0.03	0.04	0.25	0.09	-0.33	0.21	-1.36	0.09
political terror scale receiver	-0.12	0.07	-0.05	0.10	-0.07	0.21	---	---
geodistance	0.03	0.03	-0.37	0.09	0.50	0.56	-0.07	0.07

Table 7. Parameter estimates and standard errors yielded by the GERGMs fit to the four networks.

The plots below are visual representations of the parameter estimates and credible intervals for each of the four networks.

Syria



Figure 11. Plot of parameter estimates and credible intervals yielded by the GERGM fit to the Syria network.

Ukraine

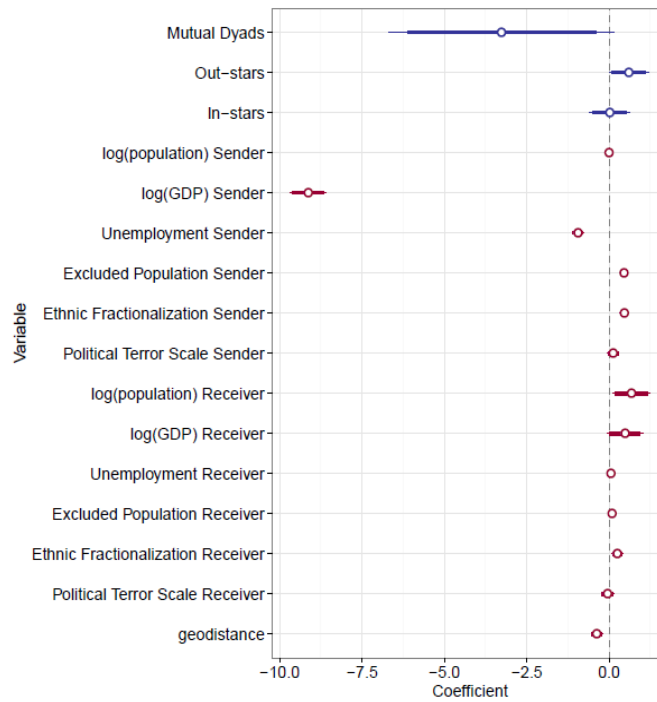


Figure 12. Plot of parameter estimates and credible intervals yielded by the GERGM fit to the Ukraine network.

DRC

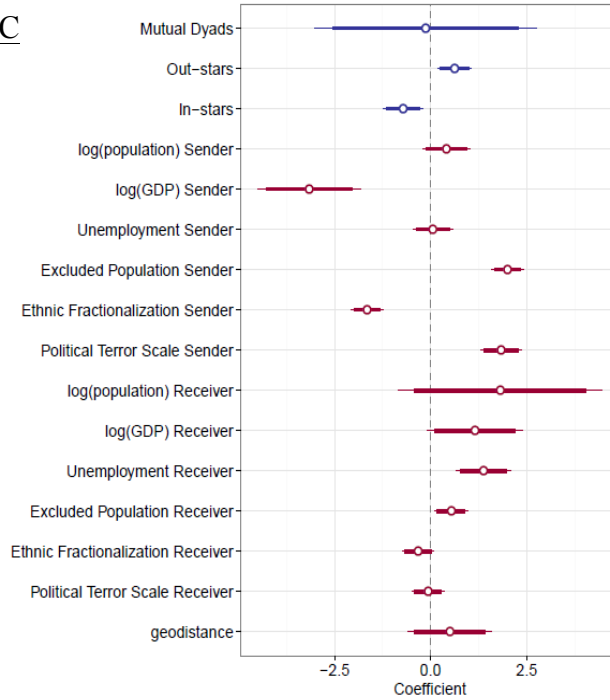


Figure 13. Plot of parameter estimates and credible intervals yielded by the GERGM fit to the DRC network.

Myanmar

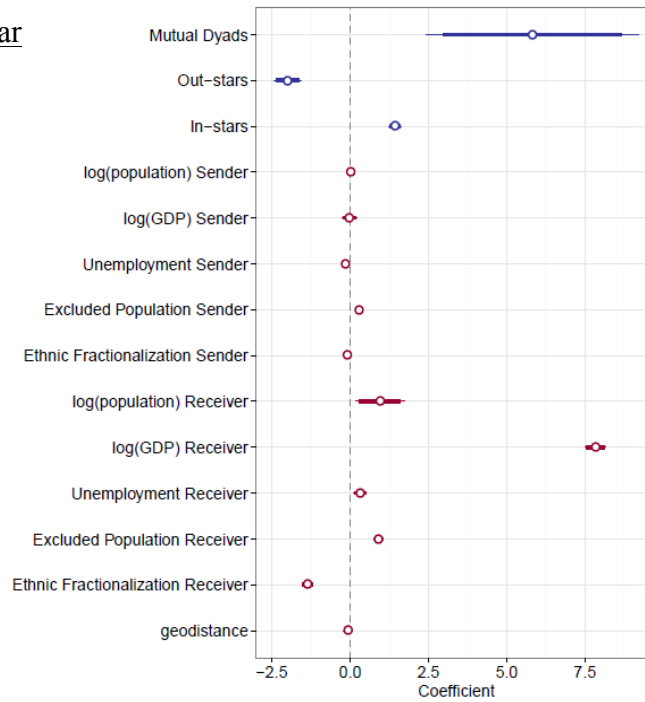


Figure 14. Plot of parameter estimates and credible intervals yielded by the GERGM fit to the Myanmar network.

If we consider the sender and receiver variables separately, we can see that the Syria and Ukraine networks are dominated by sender effects; that is, they are more significant compared to the receiver effects. On the other hand, the Myanmar network appears to be dominated by receiver effects. The coefficients estimated for the DRC network have a great amount of variation. The credible intervals are very wide. If they were slightly narrower, then there would be many significant sender and receiver effects for the DRC network.

	Syria	Ukraine	DRC	Myanmar
intercept	Red	Red	Red	Red
Mutual	Green	Red	White	Green
Out-stars	Green	Light Green	Green	Red
In-stars	Red	White	Red	Green
log population sender	Green	White	White	White
log GDP sender	Red	Red	Red	White
unemployment sender	Green	Red	White	Red
excluded population sender	Green	Green	Green	Green
ethnic fractionalization sender	Green	Green	Red	Red
political terror scale sender	Red	White	Green	White
log population receiver	Green	Light Green	White	Light Green
log GDP receiver	White	Light Green	Light Green	Green
unemployment receiver	White	White	Green	Green
excluded population receiver	White	Light Green	Light Green	Green
ethnic fractionalization receiver	White	Green	White	Red
political terror scale receiver	Red	White	White	White
geodistance	White	Red	White	White

Table 8. Summary of significance of parameter estimates yielded by the GERGMs fit to the four networks.

In Table 8, as with the ERGM, the significance of the various parameter estimates are represented by different colors in the table. White rectangles represent non-significant coefficients, while red rectangles represent negative coefficient values and green rectangles represent positive coefficient values. The lightest, middle, and darkest shades represents coefficients that are significant at the 10%, 5%, and 1% levels respectively.

Compared to the ERGMs, the GERGMs yielded many more significant parameter estimates. In terms of the structural components (mutual, in-stars, and out-stars), there are no obvious patterns. While we expected the mutual parameter (reciprocity) to be negative, it was only negative for the Ukraine network while it was positive for the Syria and Myanmar networks and for the DRC network it was insignificant. The out-stars parameter estimate was positive for the Syria, Ukraine, and DRC networks, which we would expect since this network structure represents “sociality”, or in other words, the tendency for some countries to have more out-migration than others. The networks were constructed in a way that reflects just this. Myanmar, however, had a negative parameter estimate for out-stars. The in-stars parameter estimate was negative for the Syria and DRC networks, positive for the Myanmar network, and insignificant for the Ukraine network.

Regarding the node attributes, there are some clear patterns in the results. The parameter estimates for log GDP sender are all significant and positive for all networks except for Myanmar while the parameter estimates

for log GDP receiver are all significant and negative for all networks except for Syria. Once again, this is consistent with what we would expect. The excluded population sender parameter was also strongly significant and positive for all networks, suggesting that a higher percentage of ethnically excluded population is related to higher out-migration. The log population coefficients were positive and significant for all the Ukraine, DRC, and Myanmar networks, which is consistent with the classic gravity model theory that migration is directly related to population. The coefficients for excluded population receiver were significant and positive for all networks except for Syria, which is somewhat contrary to what is suggested by the parameter estimates for excluded population sender, which are also all positive. We would expect them to have opposite signs.

Also surprisingly, only the Ukraine network yielded a significant parameter estimate for geodistance. This coefficient is negative as expected, reflecting more migration between closer countries, however it is surprising that more coefficients were not significant, and furthermore, some estimates are positive.

The diagnostics plots for each of the networks are shown in Figure 15. The plots show the network statistics calculated on the observed network and the expected values of those statistics across simulated networks. We see that they are very close, following from the first assumption of the GERGM. This indicates a good fit.

The implications of these results, as well as limitations and directions for future research, are discussed in the next section.

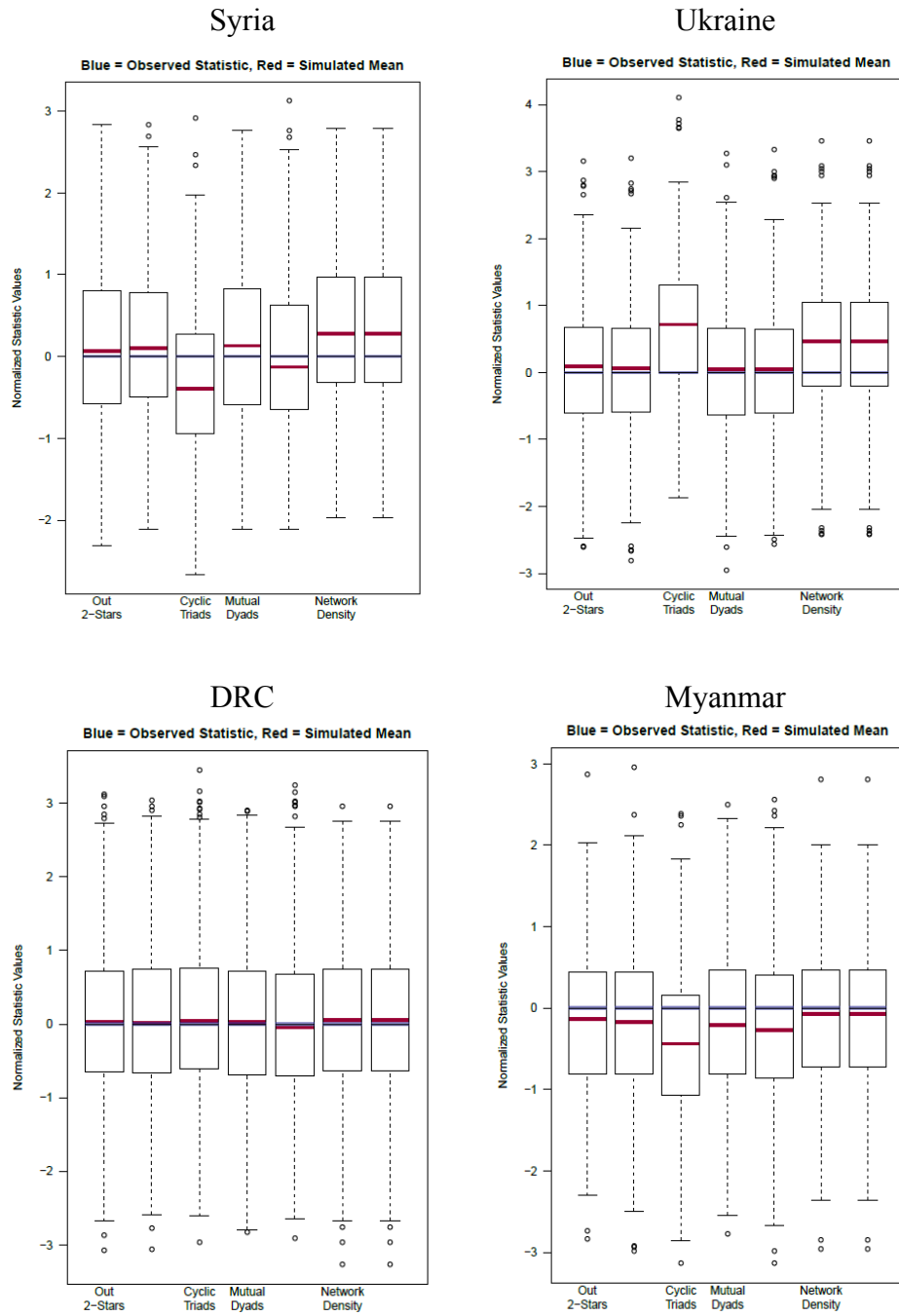


Figure 15. Diagnostics plots yielded by the GERGMs fit to each of the four networks. Clockwise from top left: Syria, Ukraine, Myanmar, DRC.

CHAPTER 5

CONCLUSION

Implications of the results

The two goals of this study were to investigate specific determinants of refugee migration for specific countries and conflicts, and to assess the strengths and weaknesses of using the ERGM and GERGM for this problem.

Perhaps the clearest conclusion that can be drawn from our results is that the ERGM is not a good model to use for this type of problem. Perhaps it would have been a feasible model if the original edge weights had not been so skewed and if the networks had been denser, however in the case of the networks involved in this study, almost all the important information about the networks was lost by thresholding. Therefore, we can conclude that for less dense networks with highly skewed weighted edges, the GERGM is the obvious choice. The GERGM is advantageous because it is flexible, allows for the inclusion of many different variables, and is easily interpreted. It does have its weaknesses, however. It does not work well with sparse networks. Also, larger networks with many nodes are increasingly difficult to model with the GERGM as the estimation process becomes extremely complicated, not to mention time-consuming. We attempted to fit many networks with 15 or more nodes and many of them took over a week to converge and most did not converge at all. Larger

networks converged when we left out some of the variables. In this way, when using the GERGM, the researcher may have to make a tradeoff between the number of nodes to include in the network and the number of variables to include.

Regarding the determinants of migration that we investigated, the results are very intriguing. Interestingly, the Syria and Ukraine networks were dominated by sender effects. There were more significant sender coefficients than receiver coefficients for these networks. This suggests that the factors we investigated influence where refugees come from in the Syrian conflict and the Ukrainian conflict, but not so much where they go. We speculate that this may be due to the fact that the conflicts in Syria and Ukraine are relatively recent – they began in the last 7 years or so. It is likely that Syrian and Ukrainian refugees are fleeing to wherever they can in order to find safety without much consideration for the economic, political, and social situation that they will find upon arrival. This may be a reason why the results of these networks were dominated by sender effects.

Surprisingly, the conflict in Myanmar is also a relatively recent conflict, but the results of this network are dominated by receiver effects. It is difficult to speculate why this might be, however one explanation might be that the vast majority of refugees are going to just a few countries which have similar characteristics in terms of the node attributes we included in the model, leading to significant receiver effects.

The results of the DRC network are very interesting. There is a great amount of variation among the parameter estimates, as indicated by the wide credible intervals, which in turn led many variables to be insignificant. It is unclear why these estimates have such wide credible intervals. Some of the edge attributes for DRC had a lot of variation, especially for excluded population, and ethnic fractionalization, but the same is true for Myanmar and we did not see similar results. Further investigation is needed in order to understand why there is so much variation in these results. However, if these intervals were slightly narrower, then many of the parameter estimates would have been significant because there are many non-zero parameter estimates for both sender and receiver effects, and the results would have been suggestive. In contrast to the conflicts in Syria and Ukraine, the conflict in DRC has been ongoing for over 20 years. In this time, an NGO presence in DRC and in neighboring countries has been established. It is possible that the parameter estimates for both sender and receiver variables are quite far from zero (while not all significant) because these NGOs have worked to put in migration routes for refugees. This might explain so many receiver effects, because the refugees have fled in an organized fashion, in contrast to the results for Syria and Ukraine. If we were to repeat this study exploring different conflicts and countries, we would expect networks built around countries with recent conflicts to display sender effects and networks built around countries with less recent conflicts to display both sender and receiver effects.

Limitations

In this section we discuss some of the limitations to our research. In terms of the data, we did perform multiple imputation on the missing values and so it is important to keep in mind that this may have had some impact on the results, especially since the multivariate normality and missing at random assumptions were violated. It is also important to keep in mind that there is the possibility that some of the data we used were not recorded correctly. This is a very common issue, especially with developing countries, especially with migration data since the chaos of the incoming refugees can make it difficult to record data accurately.

Also, the GERGM depends on the assumption that the model is specified correctly. This means that it includes all variables that are thought to have an influence on the dependent variable. Since forced migration is an incredibly complicated problem, there are certainly many variables that were not included in the model that most likely influence the dependent variable. Some variables are difficult to measure, such as migration policy, migrant preferences, and attitudes of receiving countries. While these are factors that may have a profound impact on migration flows, they cannot be included in the model if they cannot be measured. For other variables, the data are difficult to obtain or they have too many missing values.

There were many variables that we considered including in our analysis but that we left out for various reasons. One variable we left out was IAC: the presence or absence of an internally armed conflict in a given

country, a binary variable. In the ERGM and GERGM, binary variables can be included by adding a node-mixing term, which measures the extent to which nodes of the same type mix with each other. Since our networks were very small, we encountered the problem that some networks had as little as 3 successes (countries that had internally armed conflicts in 2015), which led many of our models to be degenerate. Therefore, we decided not to include a binary variable.

Other variables were left out because they were highly correlated with many of the other variables that we wanted to include. In particular, we would have liked to include polity 2, a scaled categorical variable that measures regime type, especially since many studies have found a relationship between regime type and armed conflict (Hegre, 2014). However it was highly correlated with excluded population, log GDP, and political terror scale, so we had to exclude it from our analysis. The problem of multicollinearity is a significant challenge in choosing the variables to include in the model. We attempted applying principle component analysis to try to reduce some of this multicollinearity whilst not having to exclude variables that we considered important, however we found that the resulting principle components were not interpretable, since they were not clearly dominated by any of the original variables.

Another limitation of the study is the bias that is introduced by choosing the countries to include in the network. As we mentioned earlier, the GERGM does not work well with too many nodes, and therefore we

must choose the countries in a systematic way. We chose them based on the countries that were hosting the most refugees from the respective countries of interest, however this introduces bias into the model because it completely fails to consider countries that are not taking in many refugees. Therefore we were not able to investigate why some countries are not taking in refugees, which is an equally important question. We considered other ways of constructing the networks as well in order to include countries that host less refugees. Initially, we considered creating a global network of migration including almost all countries in the world, but the network was too large to run an analysis. Also, since migration patterns differ greatly in different countries and regions, creating a global network of migration would fail to consider these differences. We also explored creating networks of countries in specific regions, choosing a country of interest and including its neighboring countries; for instance, creating a network including Syria and the 11 countries closest to Syria. One problem with creating the networks in this way is that many neighboring countries are hosting very few refugees, and so the networks would be rather sparse, which is an issue because ERGMs and GERGMs do not work well on sparse networks.

Another limitation of this study is that the data that were used for the node attributes are all from 2015, however the number of refugees living in one country from another country in 2015 (edges) are cumulative. That means we are not sure exactly when the refugees arrived in the host

country. Therefore, we cannot make strong inferences about how the node attributes relate to the number of refugees in each country. For example, many of the refugees living in DRC arrived there many years ago, when the node attributes for the countries included in the network may have been very different. For this reason, it would be very interesting to look at migration over time, comparing different networks from different years to see how the refugee situation has evolved over time, especially in countries with conflicts that have been ongoing for many years like DRC.

Another limitation of our study is that we cannot pinpoint the countries in the networks that produced the results that we obtained. We can only infer that the patterns we saw in the results are related to the countries of interest that we chose. However, since the GERGM is applied to the entire network, it takes into consideration all edges between all nodes. That means it also takes into account the migration between countries that are not the country of interest. This makes interpreting the results in the context of specific conflicts and countries very difficult, because we do not know which countries are having the most influence on the network. The collective influence of all nodes and edges is what produces the results. Consider the Ukraine network. This network includes Russia, which also has a large outflow of refugees, so perhaps the results of the Ukraine network are just as influenced by Russian refugee flows as Ukrainian refugee flows. In the future, we may want to consider different ways of building networks so that the results may be more interpretable.

Future research

In future research, we would like to explore different countries and different conflicts. We would like to further investigate our hypothesis that refugees coming from recent conflicts flee to countries without much consideration for the characteristics of the destination country, while refugees coming from conflicts that have been ongoing for a long time flee in a systematic way based on migration routes that have been established over time by NGOs. In order to investigate this hypothesis, we would like to look at multiple countries that have had recent conflicts and multiple countries that have had conflicts for many years and compare the results of these models. We would also like to include an NGO presence variable to see if that is a determining factor. Additionally we would like to look at refugee data from the past, when older conflicts first emerged, to see if we can see a similar pattern to newer conflicts today.

We would also like to include many other variables in the model, including indicators about education, language, religion, and attitudes. These are less studied factors that may have a great influence on refugee migration, however these variables are very difficult to measure.

It may also be interesting to create of model of non-forced migration, or to consider different types of migration, for example, migration due to economic reasons as opposed to violence. By narrowing the focus on the type of migration that we investigate, it may be easier to choose the variables to include in the model accordingly. For example, for

migration related to economic reasons it makes sense to include more economic variables, while for migration related to violence, it makes sense to include more political and ethnic variables.

Another interesting direction for future research would be to create a network from the change in migration from one year to another to capture the increase or decrease from year to year. We could then look at what factors we think may have caused such an increase or decrease. We could look at multiple years to paint a picture about how world events influence the increase or decrease in migration to and from certain countries and at what rate and magnitude.

Finally, in future research, we would like to explore the predictive capabilities of the GERGM. An interesting feature of the ERGM and the GERGM is that, once fitted from the data, they can predict edges. For example, we could run the GERGM for 2015 as we did. Then, once the results are obtained, we could input data from 2016 for the node and edge attributes, and based on the results from the 2015 model (parameter estimates), the model can produce estimated edge widths for 2016. This is a very interesting feature of the ERGM and GERGM that is worth exploring in the future, as it may pave the way for effective predictive models.

BIBLIOGRAPHY

- Akee, R.K.Q., Basu, A.K., Chau, N.H., Khamis, M. (2010). "Ethnic Fragmentation, Conflict, Displaced Persons and Human Trafficking: An Empirical Analysis." *IZA Discussion Paper 5142, Institute for the Study of Labor (IZA)*.
- Altai Consulting. "Migration trends across the Mediterranean: Connecting the dots." *IOM MENA Regional Office*. June 2015.
- Collett, E. "The Asylum Crisis in Europe: Designed Dysfunction." *Migration Policy Institute*. September 2015.
- Cranmer, S.J., Desmarais, B.A. (2011). "Inferential Network Analysis with Exponential Random Graph Models." *Political Analysis*, 19(1):66–86.
- Cranmer, S.J., Desmarais, B.A., Menninga, E.J. (2012). "Complex Dependencies in the Alliance Network." *Conflict Management and Peace Science*, 29 (3): 279–313.
- Crawley, H. "Chance or choice? Understanding why asylum seekers come to the UK." *Swansea University Prifysgol Abertawe and the Refugee Council*. January 2010.
- Day, K., White, P. (2001). "Choice or circumstance: The UK as the location of asylum applications by Bosnian and Somali refugees." *GeoJournal*, 55, 15-26.

- de Haas, H. (2011). "The Determinants of International Migration: Conceptualizing Policy, Origin and Destination Effects." *IMI Working Paper No 32. Oxford: International Migration Institute.*
- Dedeoğlu, D., Deniz Genç, H. (2017). "Turkish migration to Europe: a modified gravity model analysis." *IZA Journal of Development and Migration*, 7:17.
- Denny, M. (2016). "The Importance of Generative Models for Assessing Network Structure." *Pennsylvania State University.*
- Desmarais, B.A., Cranmer, S.J. (2012). "Statistical Inference for Valued-Edge Networks: The Generalized Exponential Random Graph Model." *PLoS ONE*, 7(1): e30136.
- Desmarais, B.A., Cranmer, S.J. (2017). "Statistical inference in political networks research." *The Oxford Handbook of Political Networks*. Oxford University Press.
- Frank, O., Strauss, D. (1986). "Markov graphs." *Journal of the American Statistical Association*, 81: 832–842.
- Geyer, C.J., Thompson, E.A. (1992). "Constrained monte carlo maximum likelihood for dependent data." *Journal of the Royal Statistical Society Series B (Methodological)*, 54: 657–699.
- Global Conflict Tracker. Council on Foreign Relations. Web. March 2018. <https://www.cfr.org/interactives/global-conflict-tracker#!/global-conflict-tracker>

- Görlach, J.S., Motz, N. (2017). “Refuge and Refugee Migration: How Much of a Pull Factor Are Recognition Rates?” *Bocconi University and Universidad Carlos III de Madrid*.
- Greenwood, M. (2005). “Modeling Migration.” *Encyclopedia of Social Management*, Volume 2.
- Hegre, H. (2014). “Democracy and armed conflict.” *Journal of Peace Research*, 51(2), 159-172.
- 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.
- Iqbal, Z. (2007). “The geo-politics of forced migration in Africa, 1992–2001.” *Conflict Management and Peace Science*, 24, 105–119.
- Karemera, D., Iwuagwu Oguledo, V., Davis, B. (2010). “A gravity model analysis of international migration to North America.” *Applied Economics*, 1745-1755.
- Kolaczyk E.D. “Statistical Analysis of Network Data.” New York: Springer, 2009. Print.
- Langley, S., Vanore, M., Siegel, M., Roosen, I., Rango, M., Leonardelli, I., and Laczko, F. “The Push and Pull Factors of Asylum Related Migration: A Literature Review.” *European Asylum Support Office*. November 2016.
- Lee, E.S. (1966). “A theory of migration.” *Demography*, 3(1), pp. 47-57.

- Leifeld, P., Cranmer, S.J. (2014). "A Theoretical and Empirical Comparison of the Temporal Exponential Random Graph Model and the Stochastic Actor-Oriented Model." *The 7th Political Networks Conference, McGill University*. May 2014.
- Lewis, A. (1954). "Economic development with unlimited supplies of labor." *The Manchester School of Economic and Social Studies*, 22, 139-191.
- Masri, S., Srour, I. (2014). "Assessment of the impact of Syrian refugees in Lebanon and their employment profile." *International Labour Organization, Regional Office for the Arab States*.
- Moore, W.H., Shellman, S.M. (2004). "Fear of Persecution Forced Migration, 1952-1995." *Journal of Conflict Resolution*, 48(5), 723-745.
- Neumayer, E. (2005). "Bogus refugees? The determinants of asylum migration to Western Europe." *International Studies Quarterly*, 49(3), 389-410.
- Podobnik, B., Jusup, M., Kovac, D., Stanley, H.E. (2017). "Predicting the Rise of EU Right-Wing Populism in Response to Unbalanced Immigration." *Complexity*, ID: 1580526, 12 pages.
- Poot, J., Alimi, O., Cameron, M.P., Maré, D.C. (2016). "The Gravity Model of Migration: The Successful Comeback of an Ageing

- Superstar in Regional Science.” *Institute for the Study of Labor*.
Discussion Paper No. 10329.
- Ramos, R. (2016). “Gravity models: A tool for migration analysis.” *IZA World of Labor*. 239-239.
- Ramos, R., Suriñach, J. (2013). “A gravity model of migration between ENC and EU.” *Research Institute of Applied Economics*.
- Robinson, V., Segrott, J. (2002). “Understanding the decision-making of asylum-seekers.” London: Home Office Research Study 243.
- Rother, B., Pierre, G., Lombardo, D., Herrala, R., Toffano, P., Roos, R., Auclair, G., Manasseh, K. “The Economic Impact of Conflicts and the Refugee Crisis in the Middle East and North Africa.” *International Monetary Fund*. September 2016.
- Schaeffer, P. (2010). “Refugees: on the economics of political migration.” *International Migration*, 48(1), 1-22.
- Thielemann, E. R. (2006). “The effectiveness of governments’ attempts to control unwanted migration.” *Immigration and the transformation of Europe*, 442–472.
- UNHCR. The UN Refugee Agency. Figures at a Glance. Web. March 2018. <http://www.unhcr.org/figures-at-a-glance.html>
- UNHCR. The UN Refugee Agency. Population Statistics. Web. March 2018. <http://popstats.unhcr.org/en/overview>

- Wasserman, S., Pattison, P. (1996). “Logit models and logistic regressions for social networks: I. an introduction to markov graphs and p^* .” *Psychometrika* 61: 401–425.S.
- Wilson, J., Denny, M., Bhamidi, S., Cranmer, S.J., Desmarais, B.A. (2016). “Stochastic Weighted Graphs: Flexible Model Specification and Simulation.” *Social Networks*, 49.
- Windzio, M. (2017). “The network of global migration 1990–2013 Using ERGMs to test theories of migration between countries.” *Social Networks*. SON-1045.