DYNAMIC NETWORKS OF VIOLENCE: PEOPLE POWER AND THE MEXICAN CRIMINAL CONFLICT

CASSY DORFF AND SHAHRYAR MINHAS

ABSTRACT. This study presents an aggregated analysis on the evolution of armed conflict in Mexico. The criminal war in Mexico is extremely complex: Drug Trafficking Organizations vie for control against one another and against government actors. Civilians are often subjected to DTO-related violence and left unprotected by the state. This cycle of violence is characterized by a set of interlocking interactions between armed actors. Existing research, however, typically ignores the interdependencies inherent to these networks. Using a new collection of machine-coded event data, we generate conflict networks representing each year from 2004 to 2010. Importantly, we demonstrate how these networks capture the independent nature of the Mexican conflict and test an important theoretical question of how protests influence the probability of violence between armed actors within the network. Finally we employ a latent space approach to uncover previously unobservable violence between government actors and criminal groups.

Wordcount: XXXX

Date: March 5, 2016.

Introduction

Network analysis is a critical tool for capturing dependencies inherent to the development of social phenomenon. The concept of social network analysis began as early as the 1930s with Jacob Moreno's *Who Will Survive?* and has developed into a rich and diverse field. In Political Science, network analysis has now been utilized to study a diverse range of topics: trade, intergovernmental organizations, sanctions, internal conflict, political behavior.[1] Yet, unfortunately, a majority of studies in Political Science still rely on the standard dyadic framework. We argue that conflict evolution is a process conditioned on the relative effects of actors' behavior, and is thus best conceptualized via a network approach.

Sociologists have long established that to understand an actor's behavior you have to understand the context in which they operate as well as interpret their interactions with one partner in light of all interactions across all other partners. Conflict scholars are interested in questions that mirror these exact patterns, such as: which actor is driving the violence? Did a government crackdown cause anti-government coordination or chaos between armed groups? Are civilian challenges towards non-state armed actors causing an increase or decrease in violence? These questions precisely illuminate the need to apply the appropriate methodological approach to the most pressing questions of our field.

Our paper, however, is not merely a methodological exercise. We are also motivated by a rich, underdeveloped theoretical question: do nonviolent protests influence the evolution of violence between armed actors? In this paper we utilize a new database of conflictual events in Mexico. We investigate whether civilian organizing via protests can influence the conflict between armed actors. To do so, we generate a conflict network for each year between 2006-2010. In doing so we are able to map the rise and fall of key DTOs, such

[1] For examples, see **?**, **?**, **?** and for an overview of network analysis in Political Science see **?**.

as the violent Sinaloa Cartel. We also incorporate novel protest data documenting which DTOs civilians organized against. FINDING

Existing Approaches?

Theory

Empirics

The Mexican Case. The Mexican case is the key country case for this study. The internal war in Mexico is a criminal conflict, driven by territorial disputes over trafficking routes and land and collusion between government officials and Drug Trafficking Organizations (DTOs). Drug trafficking is not a new phenomenon but over the last decade it has has been at the root of a complex conflict affecting all levels of Mexican society. After the fall of the Colombian cartels in the 1990s, the landscape of violence related to drug trafficking completely shifted in Mexico as cartels gained new territorial control. Since this time, Mexican drug cartels have become the largest foreign supplier of methamphetamine and marijuana to the United States, effectively dominating the drug market. In fact, estimates claim that the drug trade employs over half a million people and generates roughly 4% of Mexico's annual GDP.[2]

Although Mexican drug cartels have controlled the drug trade for decades, it was not until the 2006 election of Felipe Calderón that drug-related violence began to soar and civilians found themselves under fire. In 2006, Calderón became president and ushered in a new policy against the cartels. With support from the United States, the Mexican government initiated a massive campaign to combat drug-related violence. Violence soared and between 2006 and 2011 and homicides nearly tripled from 10,452 to 27,213.[3] Sending armed actors into an already armed, violent, and competitive situation, Calderon's strategy became known as a failure. It did not address the fundamental needs of civilians

[2] **?**

[3] According to Mexico's National Statistics Institute (INEGI).

or establish trusted local institutions where citizens could seek support in the realms of justice and security. Instead, these policies complicated the security situation even more and created an unstable environment for reporters, government officials, and civilians.

The failure of "Calderón's War" is partially attributable to the fact that DTOs are complex, with overlapping rivalries, family histories, splintered subgroups, and territorial disputes that drive their violent methods of political action. DTOs are also engaged in extensive corruption networks across different levels of the government and throughout the Mexican territory. The influx of federal troops into areas of high criminal activity added further complication to pre-existing corruption. Because police in Mexico receive low pay (about $9,000 to $10,000 a year), their loyalty can often be bought by cartels; however, when bribery doesn't work, cartels routinely punish government officials with violence.[4] Since combat and corruption between federal troops and cartels began, over forty mayors and numerous government officials have been murdered while increasing numbers of missing persons have been reported across Mexico as a result of cartels' increasing use of kidnapping. Government corruption, civilian victimization, and a silenced media are severe problems deeply embedded in the conflict.

Because of the cartels' brutal methods of punishment and gain, journalists and other forms of citizen representatives have been hesitant to report on these events. Journalists have not only been afraid to report out of fear that they *might* be punished; in fact, they have been targeted and killed numerous times. In 2010, Carlos Santiago, an intern photographer for the Mexican newspaper El Diario, based in Ciudad Juarez, was shot and killed. This was the second journalist from El Diario to be targeted. The other was Armando Rodrguez, a writer who worked the police beat and was killed in front of his own home. Following these deaths, the newspaper's editor drafted a plea to drug traffickers asking why they were being targeted. The article was published on the front page of the

[4] **?**

paper.[5] Then, on April 28, 2012, Regina Martínez, a journalist for the national news outlet Proceso, was found dead in her home in Xalapa, Veracruz. This series of murders is indicative of a larger phenomenon across Mexico. According to the International Press Institute and the Mexican journalists' group "Periodistas de a Pie," 103 journalists have been killed between 2000-2015 and 25 have disappeared. Since 2010, Mexico has been considered as deadly for journalists as Iraq; yet, these crimes continue with impunity. The Mexican case thus presents a relevant, timely, and difficult case for measuring the evolution of nuanced relationships between different violent actors. This study describes how the investigation of these relationships is possible.

Data Challenges in The Mexican Case. The quality of data on the Mexican criminal conflict remains mixed and generally suffers from underreporting. We know that there have been several key actors in this conflict over the years, including the Gulf Cartel, Juarez Cartel, La Familia Michoacana, Los Zetas, Sinaloa Cartel, and the Tijuana Cartel. However, because Mexican drug cartels are often in conflict with one another and infiltrated by government officials, it is difficult to attribute responsibility for homicides or other violent events to one cartel or actor versus another. Although the noisiness of this data might seem daunting, it presents an opportunity for researchers to explore how they may improve data and knowledge about violent situations in contexts where it is often dangerous to do the costly on-the-ground "legwork" that is generally necessary to accrue such information.

At present, the majority of data on violence in Mexico is based on homicide rates. Homicide data is produced from four main sources: Mexico's National Institute of Statistics and Geography (INEGI), the National System of Public Security (SNSP), the Mexican Federal Government, and La Reforma. In the beginning of the conflict (typically demarcated by Calderón's assumption of the Presidency), national newspapers carried death

[5]For a full interview with the editor see **?**.

counts related to drug violence. *La Reforma* continues to maintain drug-related homicide data; however, transparency behind the methodology of this data collection remains uncertain. It is not known, for example, how the newspaper decides whether a homicide is drug-related or not. Mexico's INEGI has data based on death certificates, which allows one to acknowledge the manner of death (such as bullet wound). This data, however, is unable to attribute which homicides are linked to crime and which are unrelated. The National System of Public Security also has crime data based on local prosecutor reports, but its reliability is questionable due to the mixed incentives for governments to accurately report information. Finally, the federal government also has released data known as the "Database of Alleged Homicides Related to Organized Crime." This database has information on executions and violence against authorities. Altogether, these data present several difficulties: first, they are not updated in real-time. To better understand the heterogenous evolution of civil conflict, researchers need to be able to describe conflict dynamics as they unfold. A further, major criticism is that these data do not further our understanding about who is directly or indirectly responsible for these crimes.[6]

Acknowledging the shortfalls of pre-existing data, my analysis improves upon existing data by providing cleaned actor event data. While I can only provide a rough estimation of actors involved in each conflictual event, this is a considerable advancement from the current status quo of knowing little to no information about which actors are engaged in which events in Mexico.[7]

ICEWS Data and The Mexican Criminal Conflict. The primary motivation for this study is to leverage machine-coded reports to construct a network of armed actors that represents conflict over time in Mexico. In order to construct this network study, I use

[6]This information summarizes an article will fuller details on the subject in *Letras Libras*. See **?**. Melissa Dell uses this government data to assess whether or not PAN victories divert drug traffic to alternative routes predicted by the shortest paths in a networked trafficking model. The aim of **?**, however, is not to create actor-based event networks.

[7]The only other data similar to this format is from a project of machine-coded data of spanish newspapers created by Javier Osorio and Alejandro Reyes. This data is not yet publicly available.

the ICEWS actor-coded event data. The ICEWS event data is part of a larger project designed to operate as a crisis warning system for policymakers.[8] This database has enabled policymakers and researchers to forecast conflictual events around the world.[9] The machine-coded event data are gleaned from natural language processing of a continuously updated harvest of news stories, primarily taken from Factiva^{TM}, an open source archive of news stories from over 200 sources around the world. The baseline event coder is called JABARI, a java variant of TABARI (Text Analysis By Augmented Replacement Instructions) which has been developed by Philip Schrodt and colleagues.[10] This approach combines a "shallow parsing" technology of prior coders with a richer exploitation of syntactic structure.[11]

The models create each data point by obtaining three components of the news story: the sender of the event (i.e., who initiated the action), the receiver or target of this action, and then the event type itself. I subsetted this data according to relevant "violent" cameo codes in order to gain access to all events relating to any armed actors such as rebels, insurgents, government, and the police. These events, in essence, capture any type of violent conflict between different actors. The event type itself is coded according to the Conflict and Mediation Event Observation (CAMEO) ontology.[12] The main distinguishing feature of CAMEO is its use of mediation related event codes. CAMEO does not assume that a meeting is a peaceful interaction, for example, but is able to decipher whether meetings between actors are related to mediation, or negotiation. CAMEO also includes four categories for violence (structural violence, unconventional violence, conventional force, and massive unconventional force) as well as a rich system of sub-categories.

[8]For a summary, see **?**.

[9]**?**

[10](see http://eventdata.psu.edu/)

[11]This has increased accuracy (precision) from 50% to over 70%, as demonstrated in a series of ongoing (informal) evaluations of its output by human graders. Peak human coding performance is reported to be around 80% (**?**).

[12]See **?** for the full summary of the project.

To begin to understand how to leverage the ICEWS data for country-level network analysis, I have taken a subset of data from the larger ICEWS corpus. I constructed a SQL query to gain data subsetted according to all four "violent" cameo codes as well as any actions related to all armed actors such as rebels, insurgents, government, and the police. Through the process of reviewing and cleaning the ICEWS data in preparation for my analysis, I've encountered two key problems with the data. The first problem relates to the vague nature of the actor names in the data, which I improve upon via manual re-coding. The second problem I identify incentivizes ICEWS programmers to improve the parsing algorithm used in the creation of the original data. This original raw, deduplicated, data from 2004-2010 contains around 200 observations relating to events between armed actors. While there are many unique actor names, the bulk of these descriptions are likely too vague for network construction. For example, the majority of cases relating to criminal violence use descriptors such as: "Armed Gang," "Armed Opposition," "Attacker," "Hitman," "Drug Gang," "Armed Band," and "Criminal." A number of other cases have actor names such as "Men" or "Citizens" as well as Military descriptors.

To improve on these issues, I subset the data to include the raw text available from the larger ICEWS database. Then, using a subset of cases from 2004-2010, I reviewed each individual case. This task includes two main goals: to label events as they relate to specific drug cartel actors in the area and to address aggregation problems in the data. In addition, I coded for duplicate cases. I identified that this data has fewer duplication issues than previously found in other ICEWS data (such as protests) but has a number of complex aggregation and parsing problems.

In addition to correcting for the number of conflictual events over time, I also correct for the vague coding descriptors found in the original data. To do this, I read the stories and coded whether a specific criminal group, cartel, or cartel member was mentioned. I then code the new actors' names and list any relevant actor involved in the conflictual event. A variety of other news sources, blogs, and area expert knowledge were used to complete this new set of actor codings. If I could not locate any resources that allowed me to identify which cartel, or actor, was involved in the event, I utilized a pre-existing data set containing the locations of cartels over time. This data set was created by Viridiana Rios and Michele Coscia and records locations of cartels down to the municipal level. An example of the territorial changes for the Sinaloa Cartel are shown in Figure 1.[13] Exploiting online newspapers and blogs, they develop a mechanism that uses unambiguous query terms to classify the areas in which criminal organizations operate.

[13]Maps and data provided by Viridiana Rios, which can be accessed online, see **?**.

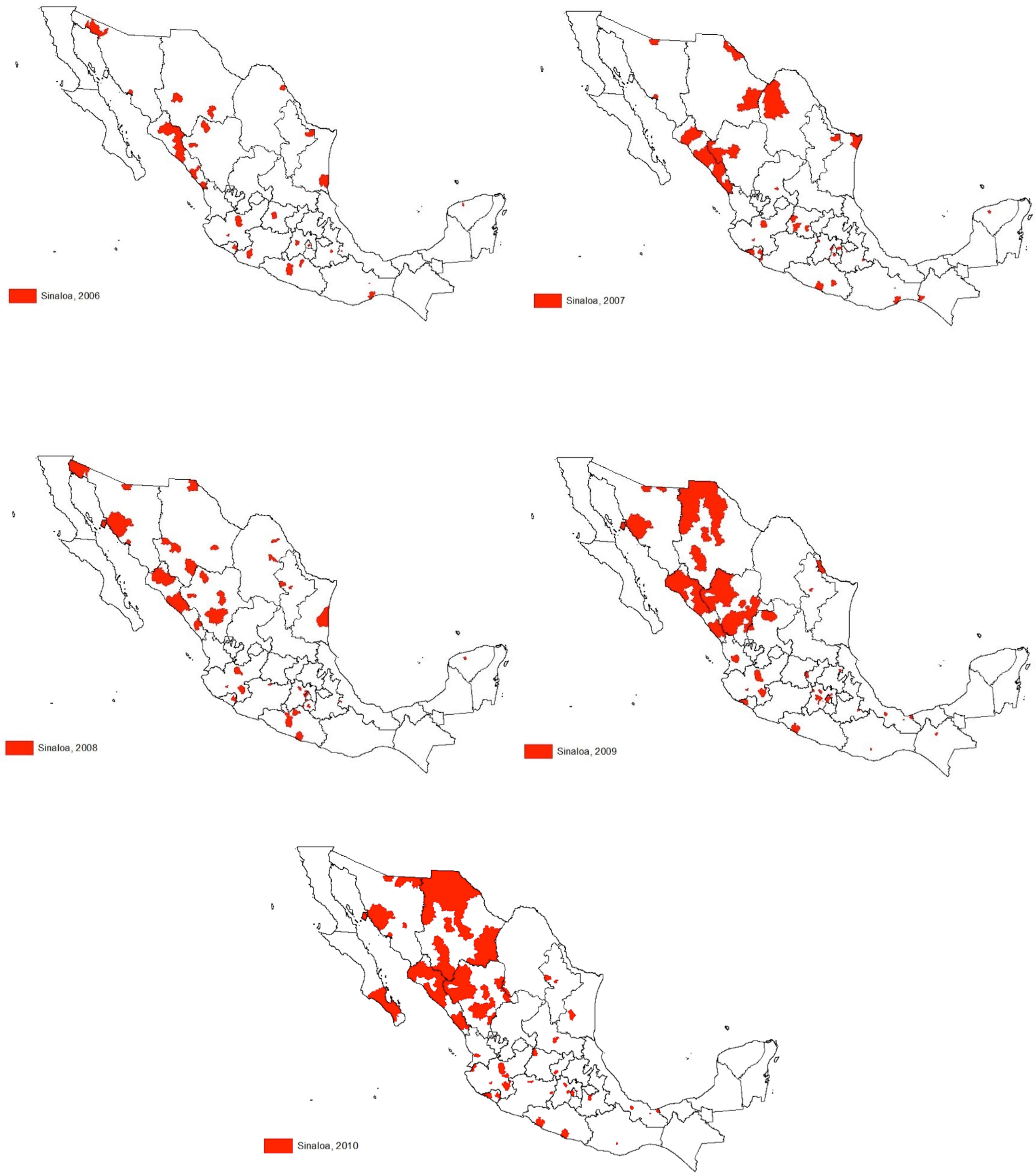

Figure 1. Sinaloa Cartel's territorial movements from 2006-2010

WHAT THE DATA SHOWS (COUNTS/TRENDS), UPDATE ACTOR LIST The trajectory of the data between 2008 and 2009, however, differs significantly. In the original data it would seem that during this time violence is only slightly higher than before the

2007 spike. With the new data we are able to see that violence is much higher than the pre-2007 levels and seems to be following an increasing pattern. Perhaps most importantly, this new data allows for the creation of network data that allows future research to explore how other political events of interest influence the evolution of this network.

Table 1. List of updated actor names 2004-2010

	Actor Names
Army	13
Beltran-Leyva Cartel	12
Business	1
Citizen	62
City government	7
Federal government	2
Federal police	35
Gulf cartel	13
Indigenous	2
Juarez cartel	12
La Familia	15
Los Zetas	15
Military	9
Municipal government	6
Municipal police	40
Navy	2
Sinaloa cartel	13
Special forces	2
State government	5
State police	16
Tijuana cartel	3
Unknown	5
Total	308

Creating a Conflict Network in Mexico

In the final stage of the analysis, I create sociomatrices for each year of the cleaned data. These sociomatrices can be thought of as a summary of interactions between all actors involved in conflictual events within a year. Given that there are n actors in a year I construct an n x n sociomatrix Y. The number of conflictual dyadic interactions for any actor i and j is simply the number of events between those two actors during each given year. The resulting matrix is an undirected, symmetric matrix, as represented below.

$$\begin{bmatrix} & actor_i & actor_j & actor_k & actor_l & actor_m \\ actor_i & 0 & 2 & 0 & 0 & 0 \\ actor_j & 2 & 0 & 0 & 2 & 1 \\ actor_k & 0 & 0 & 0 & 0 & 4 \\ actor_l & 0 & 2 & 0 & 0 & 4 \\ actor_m & 0 & 1 & 4 & 4 & 0 \end{bmatrix}$$

These matrices reveal a variety of interesting dynamics in the data, as shown in the network graphs of Figure **??**, Figure **??**, and Figure **??**. First, Figure **??** shows the network graph for the first two years of the data. Beginning as early as 2004 to 2005, there is a substantial increase in the number of actors in the network. We observe that "Citizens" has a high degree in both networks or, rather, that the Citizen node is linked to a high number of other nodes within the group. In both of these networks we also see that the municipal police are the most heavily involved in these conflictual events, whereas other levels of government, such as the federal police, play a minor role in this time period.

NETWORK PICTURE HERE

These dynamics shift in 2006 and 2007, shown in Figure **??**. Here we see that both federal agencies and municipal forces are active in the conflict. Citizens are also entangled in these events, and the number of total actors increases from 12 in 2005 to 16 in 2006.

There is a strong triad forming between citizens, federal police, and municipal police. We also observe a jump in the number of cartels involved in the network, as new actors such as La Familia Michoácana, the Mexican Mafia, and the military begin to take part in the conflict. Notably, this network also shows that the federal police and municipal police are now both highly active. This reflects a change in the government policy at that time. By 2006 Calderón was elected president and implemented a militarized strategy, sending federal police and troops into the most contested regions. By 2007, the network has 18 actors and we see that the relationships become more complex as cartels begin to interact with one another. This sufficiently reflects what we now know to be true: following Calderón's 2006 efforts, violence increased due to grueling competition between newly fragmented cartels.

2nd NET PIC HERE

In 2008, similar dynamics are in motion, but it is difficult to decipher the various roles played by the different levels of government. Earlier, both federal and municipal governments seem similarly involved, but by 2008 and 2009 the municipal government has a broader connectivity to diverse actors while the federal police seems to become more limitedly engaged with the Sinaloa Cartel. Overall, density increases across these networks overtime. Finally, the years 2008 and 2009 demonstrate how often citizens are caught in the mix of the violence. A consistent link across networks is that of civilians who seem to be caught in the crosshairs of much of the violence.

To further assess how different levels of government change their role in the network over time, Figure 2 presents the eigenvalue centrality for each actor. Eigenvector centrality is calculated by assessing how well connected an actor is to the other actors of the network. Specifically, those with the highest eigenvector centrality are well-connected to other actors who are also well-connected across the network. In this case we can consider an actor with high eigenvector centrality to be a key player in the conflict.

final net pic here

Figure 2 reveals an interesting story of coordination between different levels of government. In 2004 both the municipal police and the federal police have similar levels of eigen centrality. In 2005 they both decrease, with a larger decrease for the municipal police. In 2006 we see that both actors have very high centrality, indicating a deep involvement in the conflict. This reflects the Calderón strategy: sending federal forces into local areas to help bolster security enforcement and combat the cartels. We then see these two actors diverge. Federal police stay highly central in the network in 2007 while the municipal forces seem to back down. At this year these two actors experience the biggest gap in centrality, signaling that coordination between the two levels of governments decreases as federal forces take over combat operations relative to municipal security forces.

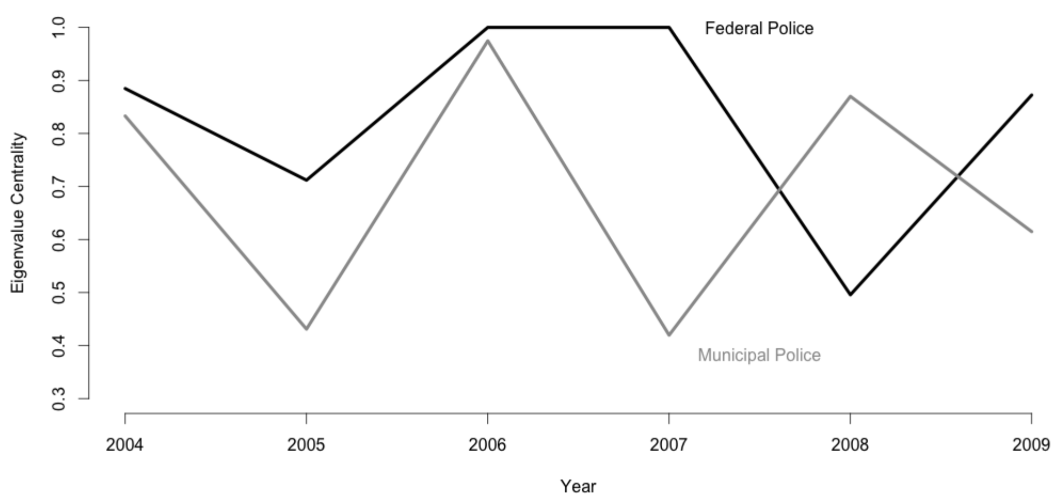

Figure 2. Eigen centrality at the yearly level (2004-2009) for both municipal police and federal police

By 2008, we see a stark shift as municipal forces are once again involved in the conflict while federal forces seem to back down. This is particularly interesting because at this time it becomes well known to the public and to the government that Calderón's strategy has largely failed, causing more violence rather than stemming it. Figure 2 shows that after federal forces are very active from 2006-2007, in 2008 they then recede and the municipal police become more active in the network. Finally, the year 2009 shows closer centrality scores than 2008 and 2007, signaling that perhaps these two levels of government begin to cooperate again.

Without the re-structuring of the event data into networks we would not be able to see the ways in which different levels of governments interact over time. This approach offers a more complete picture of these dynamics than has previously been possible and enables future research to examine what factors influence government coordination within the network.

Latent Space Analysis

A further analysis of these networks allows us to answer questions about the probability of interaction between actors in a network. While there are several approaches to social network analysis, I employ the latent space approach. Latent space approach is most useful when the main goal is to understand the role of individual actors in the network. Specifically, the latent space approach can identify interactions that are unobserved in the raw data. This approach is presented by **?** and has been used in political science in several applications.[14]

The essential idea of the latent space is to capture third-order dependence. A common example involves relationships within a triad i, j, k. If we know that i considers j a friend and j is a friend of k, then the probability that k will also be a friend of i is likely to be higher than for a random person outside of this triad, since i and k are at least indirectly connected in the friendship network by virtue of their separate linkages to j. Thus, information about the relationships in the first two dyads of a triad can usually reveal something about the relations in the third dyad. Third-order dependence, or the "unobserved," latent social space then becomes a highly useful concept. The latent space can be thought of as a probability space, whereby observation of two links, i-j and j-k,

[14]For examples regarding political conflict see **?** or **?**

suggests that i and k are not too far away from each other in this social space and therefore are also likely to have a link between them. Since third-order dependence is an expression of the underlying probability of a link between two actors, we do not observe the complete set of all of these network characteristics, but we can infer them from the pattern of dyadic linkages. If we can map out the latent positions of each actor in the "social space," we can then assume that the ties in the network are conditionally independent.

Formally, if we are interested in modeling an n x n sociomatrix that contains dyadic data, we might do so with a typical linear regression approach:

$$y_{ij} = \beta' x_{i,j} + \epsilon_{i,j}. \tag{1}$$

While this approach is certainly common, it assumes the errors, $\epsilon_{i,j}$ are independent. In employing the General Bilinear Mixed Effects modeling (GBME) approach, we alter the assumption of the errors and instead assume that the errors $\{\epsilon_{i,ji} \neq j\}$ have a covariance that is exchangeable under identical permutations of the indices i, j. We then assume normality, which implies that the residuals can be represented as a linear random-effects model with sender (a_i) and receiver (b_j), and dyadic y_{ij} effects:

$$\begin{gathered}
\epsilon_{ij} = a_i + b_j + \gamma_{i,j} \\
\begin{bmatrix} a_i \\ b_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ba} & \sigma_b^2 \end{bmatrix} \right) \\
\begin{bmatrix} \gamma_{i,j} \\ \gamma_{j,i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\gamma^2 & \rho\sigma_\gamma^2 \\ \rho\sigma_\gamma^2 & \sigma_\gamma^2 \end{bmatrix} \right).
\end{gathered} \tag{2}$$

This allows us to estimate the following moments:

$$
\begin{aligned}
E(\epsilon_{i,j}^2) &= \sigma_a^2 + \sigma_b^2 + \sigma_\gamma^2 \\
E(\epsilon_{i,j}\epsilon_{j,i}) &= \rho\sigma_\gamma^2 + 2\sigma_{ab} \\
E(\epsilon_{i,j}\epsilon_{i,k}) &= \sigma_a^2 \\
E(\epsilon_{i,j}\epsilon_{k,j}) &= \sigma_b^2 \\
E(\epsilon_{i,j}\epsilon_{k,i}) &= \sigma_{ab}.
\end{aligned} \tag{3}
$$

where σ_a^2 represents dependence among dyadic observations with a common sender, σ_b^2 represents dependence among measurements having a common receiver, and ρ is the correlation of measurements within a dyad, or *reciprocity*. To adjust for other types of data, such as the count data used here, the error structure can be altered so that the dyadic data are conditionally independent given the random effects but are unconditionally dependent:[15]

$$
\begin{aligned}
\theta_{i,j} &= \beta' x_{i,j} + a_i + b_j + \gamma_{i,j} \\
E(y_{i,j}|\theta_{i,j} &= g(\theta_{i,j}) \\
p(y_{1,2}...y_{n,n-1}|\theta_{1,2}...\theta_{n,n-1}) &= \prod_{i \neq j} p(y_{i,j}|\theta_{i,j}).
\end{aligned} \tag{4}
$$

Following **?**, we can define the unobserved, K-dimensional vector z_i for each node i in the network. By modeling the interaction of two nodes as an increasing function of their proximity in the latent space, we include patterns of transitivity, balance, and clusterability into the network. Formally, we can incorporate this into the model by adding the inner product $z_i' z_j$ to the linear predictor:

$$
\epsilon_{ij} = a_i + b_j + \gamma_{i,j} + z_i' z_j. \tag{5}
$$

[15]Where g(·) is the inverse-link function. This is a summary of the full specification provided in **?** and **?**.

Conclusion

Appendix

Cassy Dorff: Korbel School of International Studies

Current address: University of Denver, Denver, CO

E-mail address: cassy.dorff@du.edu

Shahryar Minhas: Department of Political Science

Current address: Duke University, Durham, NC, 27708, USA

E-mail address: shahryar.minhas@duke.edu