Pre-Analysis Plan: Network Theory of the Democratic Peace

September 2, 2016

Contents

1 Introduction **1**
- 1.1 Abstract . . . 1
- 1.2 Motivation . . . 1
- 1.3 Research Questions . . . 1
- 1.4 Building upon Cranmer and Desmarais (2011) . . . 1

2 Research Design **2**
- 2.1 Sampling . . . 2
- 2.2 Variable Descriptions . . . 2
  - 2.2.1 Joint Democracy (Indicator) . . . 3
  - 2.2.2 Democracy (Low) . . . 3
  - 2.2.3 Democracy (High) . . . 3
  - 2.2.4 Dyadic Difference Democracy . . . 3
  - 2.2.5 Joint Autocracy (Indicator) . . . 3
  - 2.2.6 Mixed Pair (Indicator) . . . 3
  - 2.2.7 Proportion of Democracy . . . 3
  - 2.2.8 Years of Peace, Years of Peace Splines . . . 4
  - 2.2.9 Number of States in the International System . . . 4
  - 2.2.10 Capital-to-Capital Distance . . . 4
  - 2.2.11 Contiguity . . . 4
  - 2.2.12 Alliance . . . 4
  - 2.2.13 Capability Ratio . . . 4
  - 2.2.14 Major Power . . . 5
  - 2.2.15 CINC Score of Strongest State . . . 5
  - 2.2.16 2-Stars . . . 5
  - 2.2.17 Triangles . . . 5
- 2.3 Data Access and Processing . . . 5
- 2.4 The Bootstrapped Temporal Exponential Random Graph Model (BTERGM) . . . 8
- 2.5 Interpreting Results . . . 10

3 Auxilliary Analyses **10**
3.1 Distinct Measures of Democracy 11
3.2 Model Estimating using Fatal MIDs 18
3.3 Examining the Dyads that Contribute Most to the Democratic Peace 21
3.4 Examining Temporal Heterogeneity in the Democratic Peace 23
3.5 Examining All Participants 24

PRELIMINARY DRAFT: PLEASE DO NOT CITE OR DISTRIBUTE

1 Introduction

1.1 Abstract

This pre-analysis plan outlines a research strategy for testing a network-topological theory of the democratic peace. The democratic peace always has been, and continues to be, considered a dyadic phenomenon explained through theories with dyadic implications. Our theory holds that the democratic peace is less a result of democracies' unwillingness to fight one another, but more representative of a network process where democracies have common enemies, and as a result, do not fight one another. These shared enemies emerge from coordinated responses to international emergencies and conflicts. We expect that states fighting a common enemy would not fight one another as it would be inefficient and to do so – the enemy of my enemy is not my enemy. This theory will be tested on a dataset compiled by Gartzke and Weisiger (2013) and cleaned by Dafoe, Oneal and Russett (2013) using the Temporal Exponential Random Graph Model as described by Cranmer and Desmarais (2011).

1.2 Motivation

The Democratic Peace is one of the foundational findings in IR, acquiring a law-like status for some (Levy 1998; Hegre 2014). This finding is considered to be quite robust, and can be reduced as follows: jointly democratic states do not go to war. While many have challenged the democratic peace using conventional large-N covariates, few have used the tools of inferential network analysis to challenge its fundamental premise (Cranmer and Desmarais 2011). We attempt to correct this through testing a networks-based explanation of the democratic peace.

1.3 Research Questions

Is there a network-based explanation for the democratic peace? Is the democratic peace a function of shared enemies between democracies? Are democracies less likely to fight each other because of regime type, or inefficiencies created by triadic closure in the conflict network?

1.4 Building upon Cranmer and Desmarais (2011)

Cranmer and Desmarais (2011) challenge the fundamental theoretical implications of the democratic peace (as represented by Maoz et al. (2006)), arguing that two features of the conflict network may explain these results (81). First, a state's "popularity" in the conflict network, or the degree of targeting they experience, may explain certain dynamics that covariates may not capture. In particular, this popularity term would capture organized and internationally coordinated responses to diplomatic and security crises. This term is captured in the network topological context using a two-star statistic, which in the context of the conflict network would reference the number of times in the network where two states are at war with the same state (Cranmer and Desmarais 2011). Second, Cranmer and Desmarais (2011) specify a "triangle" network statistic, which they argue should not be present relative to the two-star statistic. The triangle statistic refers to the count of fully completed 3-node cycles. In this particular case, it refers to the number of instances where state i is at war with j, who is also engaged in conflict with state k, who is at war with state i. In this particular case, it would make very little sense for i and kk to go to war as it would undermine their mutual effort against j – we would expect the enemy of a state's enemy to be its friend.

Our work builds upon the findings of Cranmer and Desmarais (2011), presenting a well-developed network theory of conflict that would account for much of the democratic peace, and a thorough test of that theory using the Temporal Exponential Random Graph Model (TERGM). The TERGM allows scholars to test network-based theories on longitudinally observed networks and provides the perfect method for examining something as dynamic as the international conflict network.

2 Research Design

This section describes the research design used in our primary analysis. In particular, we describe the initial source for our data, how we may manipulate it, and the model used to test our network theory. This section concludes with outlining the conditions that must be met for us to find resolute support for our theory.

2.1 Sampling

Our sample draws from the data initially sourced from Gartzke and Weisiger (2013) and cleaned by Dafoe, Oneal and Russett (2013). The dataset is constituted by $656,317$ dyad-years. These reflect the population of dyads available for each given year between 1816 and 2001. For a more complete understanding of the population of cases, see Gartzke and Weisiger (2013). The data that we draw from, Dafoe, Oneal and Russett (2013), makes the following modifications to the Gartzke and Weisiger (2013) dataset:

- In some periods Gartzke and Weisiger (2013) includes microstates where there is significant missingness on relevant covariates, including POLITY IV. These microstates were initially included in Gartzke and Weisiger (2013) as missing values coded as an active POLITY scale value (coded as -9), and were correctly cleaned by Dafoe, Oneal and Russett (2013) to treat them as missing (NA).
- There are also coding problems in the contiguity variable. Gartzke and Weisiger (2013) use the COW *colcont* cariable which only captures states that are contiguous through colonial possession-to-state shared borders, which is to say, it does not include state-to-state shared borders (Dafoe, Oneal and Russett 2013). An example where this is problematic in the Gartzke and Weisiger (2013) data is the case of Canada and the United States. Canada and the United States are only contiguous until 1959 when Alaska is a colonial possession of the United States, they are not contiguous once Alaska becomes a state.
- The major power variable, which is a binary indicator in Gartzke and Weisiger (2013) for if one state in the dyad is a major power, is replaced with a continuous measure of the size of the most powerful state in the dyad (CINC scores) (Dafoe, Oneal and Russett 2013).
- Dafoe, Oneal and Russett (2013) add a scaling term that accounts for the number of states in the system to account for the number of opportunities for conflict.

2.2 Variable Descriptions

In this subsection, we describe the variables that will be included in two different model specifications. We plan on fitting two models, one based upon the Gartzke and Weisiger (2013) specification

and one based upon the Dafoe, Oneal and Russett (2013) specification.

2.2.1 Joint Democracy (Indicator)

The indicator for a jointly democratic dyad is used in the conditional effects calculated by Dafoe, Oneal and Russett (2013). This variable takes on a value of one for the dyad if the state with the lowest polity value has a value greater than 6.

2.2.2 Democracy (Low)

This variable, used by (Gartzke and Weisiger 2013), takes on a value equal to the lowest polity value for a dyad.

2.2.3 Democracy (High)

This variable, used by (Gartzke and Weisiger 2013), takes on a value equal to the highest polity value for a dyad.

2.2.4 Dyadic Difference Democracy

Dyadic difference is an important variable used in the conditional effects calculated by Gartzke and Weisiger (2013) and is calculated as the difference in the highest polity value of a state in the dyad and the lowest polity value of a state in the dyad. This reflects an attribute of a dyad.

2.2.5 Joint Autocracy (Indicator)

The indicator for a jointly autocratic dyad is used in the conditional effects calculated by Dafoe, Oneal and Russett (2013). This variable takes on a value of one for the dyad if the state with the largest polity value has a value less than or equal to than 6.

2.2.6 Mixed Pair (Indicator)

The indicator is for a mixed democratic-autocratic and is used in the conditional effects calculated by Dafoe, Oneal and Russett (2013). This variable takes on a value of one for the dyad if the state with the largest polity value has a value less than or equal to one and the state with the highest polity value has a polity value greater than 6.

2.2.7 Proportion of Democracy

The proportion of democracies in the system-year variable is calculated as a year-based variable that takes on a value equal to the quotient of the number of states with polity values greater than 6 and the total number of states in the system for a given year. The theoretical intutition for including these variables is discussed at length in Dafoe, Oneal and Russett (2013) and Gartzke and Weisiger (2013).

2.2.8 Years of Peace, Years of Peace Splines

To account for temporal dependence, Dafoe, Oneal and Russett (2013) and Gartzke and Weisiger (2013) specifies two variables. The first is the peace-years for a given dyad which is defined as the number of years since the last event. Second, square and cubic splines are also specified as a function of the peace-years variable. These variables are suggested by Beck, Katz and Tucker (1998) as a way to account for the effect that time may have on event occurance.

2.2.9 Number of States in the International System

In following the advice of Raknerud and Hegre (1997), Dafoe, Oneal and Russett (2013) include a control for the number of states active in the international system that could lead to a greater prevalence of conflict. This variable is calculated as the natural log of the number of states that exist in the international system for a given year.

2.2.10 Capital-to-Capital Distance

This variable is defined as the natural log of the great circle distance between national capitals or the closest major cities for larger countries (Gartzke and Weisiger 2013). Distances between states is an important component of political relevance, and as such, it makes great sense to ensure its inclusion (Braumoeller and Carson 2011). Distance between states may also confound regime-type, as regime-types cluster, and the likelihood of conflict (Cederman and Gleditsch 2004).

2.2.11 Contiguity

Contiguity is another relevant measure of distance between states and is a typical control employed in the democratic peace literature (Gartzke and Weisiger 2013). This variable is ordinal and describes six decreasing levels of proximity, from shared land borders to separated by no more than 500 miles of water.

2.2.12 Alliance

Allies are less likely to go to war with one another, and as such, studies on the democratic peace often include an indicator variable for whether or not two states in a dyad share a joint alliance commitment (Gartzke and Weisiger 2013).

2.2.13 Capability Ratio

Studies of the democratic peace very typically include a measure of the relative capabilities of a state and the ability of states to conduct warfare. We include a variable that captures this dynamic, capability ratios, which are calculated as the the quotient of the weakest state in a dyad's CINC score and the sum of CINC scores for the dyad:

$$CR_{AB} = \frac{CINC_{low}}{CINC_A + CINC_B} \tag{1}$$

The above is how it is calculated by Gartzke and Weisiger (2013). However, Dafoe, Oneal and Russett (2013) defines the capabilities ratio as follows and refers to it as the probability of a state

with the highest capabilities winning:

$$CR_{AB} = \frac{CINC_{high}}{CINC_A + CINC_B} \tag{2}$$

2.2.14 Major Power

Gartzke and Weisiger (2013) utilize an indicator variable for if at least one state in a dyad is a major power as defined by the Correlates of War.

2.2.15 CINC Score of Strongest State

As an alternative to the binarized coding of major power status in a dyad used by Gartzke and Weisiger (2013), Dafoe, Oneal and Russett (2013) instead opt to use a continuous measure that takes on value equal to the greatest CINC score in a dyad.

2.2.16 2-Stars

We argue that the number of incoming ties a state experiences may explain certain dynamics that covariates cannot capture. In particular, this popularity term captures organized and internationally coordinated responses to diplomatic and security crises. Considering this in the context of network topology, we specify the count of two-stars that exist in a network-year. In the context of the conflict network, this would reference the number of times in the network where two states are at war with the same state (Cranmer and Desmarais 2011). So, assuming there are three states, i, j, k, a two-star would be if $i \rightarrow j$ and $k \rightarrow j$.

2.2.17 Triangles

In addition to the two-star statistic, we specify a "triangle" network statistic, which should not be present in the conflict network relative to the two-star statistic (Cranmer and Desmarais 2011). The triangle statistic refers to the count of fully completed 3-node cycles. In the conflict network, it refers to the number of instances where state i is at war with state j, who is being attacked by k, who is at war with state i. In this particular case, it would make very little sense for j and k to go to war as it would undermine their mutual effort against i – we would expect the enemy of a state's enemy to be its friend.

2.3 Data Access and Processing

Data for our analysis will be sourced from Dafoe, Oneal and Russett (2013). The following chunk of code processes the Dafoe, Oneal and Russett (2013) data so that it can be used by the the code listed in a later section:

```
# Set local working directory
setwd("~/Box Sync/NetworkDemocraticPeace")

# Load relevant packages
library(foreign)
library(statnet)
```

```r
library(data.table)
library(DataCombine)
# Load data from Dafoe, Oneal, and Russett
  # This data was prepared according to their data cleaning script
  # and then exported from Stata12
dat <- read.dta("dor_18162001_08252016.dta")

# Dyadic difference variable
dat$dyaddiff <- dat$lrgdem - dat$smldem
dat$dyadsyst <- dat$dyaddiff*dat$propdem
dat$caprat <- with(dat, smlcap/(smlcap + lrgcap))
dat$majpow <- ifelse(dat$majpow1 | dat$majpow2 == 1, 1, 0)

# Dataset for G&W replication model 4 (originators only, normal MIDs), but with data fixes
gw_vars <- c("statea", "stateb", "dyadid", "year", "mzmid1_orig", "smldem",
             "dyaddiff", "propdem", "dyadsyst", "lndstab", "contiguity",
             "allies", "caprat", "majpow", "py", "_spline1", "_spline2", "_spline3")

gw.dat <- NULL
for(i in gw_vars){
  gw.dat <- cbind(gw.dat, dat[,which(names(dat)==i)])
}

gw.dat <- as.data.frame(gw.dat)
names(gw.dat) <- gw_vars
gw.dat <- na.omit(gw.dat)

# Create triadic change stats
tchange <- numeric(nrow(gw.dat))
tschange <- numeric(nrow(gw.dat))

for(i in 1:nrow(gw.dat)){
  dat1 <- subset(gw.dat, gw.dat$year==gw.dat$year[i])
  dat1mid <- subset(dat1, dat1$mzmid1_orig==1)
  datc1 <- subset(dat1mid, as.numeric(dat1mid$statea == gw.dat$statea[i]) +
                    as.numeric(dat1mid$stateb == gw.dat$statea[i]) > 0)
  datc2 <- subset(dat1mid, as.numeric(dat1mid$statea == gw.dat$stateb[i]) +
                    as.numeric(dat1mid$stateb == gw.dat$stateb[i]) > 0)
  us1 <- unique(c(datc1$statea, datc1$stateb))
  if(is.element(gw.dat$statea[i], us1)) us1 <- us1[-which(us1 == gw.dat$statea[i])]
  if(is.element(gw.dat$stateb[i], us1)) us1 <- us1[-which(us1 == gw.dat$stateb[i])]
  us2 <- unique(c(datc2$statea, datc2$stateb))
  if(is.element(gw.dat$stateb[i], us2)) us2 <- us2[-which(us2 == gw.dat$stateb[i])]
  if(is.element(gw.dat$statea[i], us2)) us2 <- us2[-which(us2 == gw.dat$statea[i])]
  if(min(c(length(us1),length(us2)))>0){
```

```
    tchange[i] <- length(intersect(us1,us2))
    tschange[i] <- length(us1)+length(us2)
  }
}

gw.dat <- cbind(gw.dat, tchange, tschange)
save(gw.dat, "gw_dat_triadicChange.RData")

# Now, we create the dataset for the DRO dataset

dro_vars <- c("statea", "stateb", "dyadid", "year", "mzmid1_orig", "jtdem",
              "propdem", "jtdemxpropdem", "jtaut", "jtautxpropdem", "mixed",
              "mixedxpropdem", "contiguity", "lndstab", "pwin_lrg", "lrgcap",
              "allies", "lnNt", "py", "_spline1", "_spline2", "_spline3")

dro.dat <- NULL
for(i in dro_vars){
  dro.dat <- cbind(dro.dat, dat[,which(names(dat)==i)])
}

dro.dat <- as.data.frame(dro.dat)
names(dro.dat) <- dro_vars
dro.dat <- na.omit(dro.dat)

# Create triadic change stats
tchange <- numeric(nrow(dro.dat))
tschange <- numeric(nrow(dro.dat))

for(i in 1:nrow(dro.dat)){
  dat1 <- subset(dro.dat, dro.dat$year==dro.dat$year[i])
  dat1mid <- subset(dat1, dat1$mzmid1_orig==1)
  datc1 <- subset(dat1mid, as.numeric(dat1mid$statea == dro.dat$statea[i]) +
                    as.numeric(dat1mid$stateb == dro.dat$statea[i]) > 0)
  datc2 <- subset(dat1mid, as.numeric(dat1mid$statea == dro.dat$stateb[i]) +
                    as.numeric(dat1mid$stateb == dro.dat$stateb[i]) > 0)
  us1 <- unique(c(datc1$statea, datc1$stateb))
  if(is.element(dro.dat$statea[i], us1)) us1 <- us1[-which(us1 == dro.dat$statea[i])]
  if(is.element(dro.dat$stateb[i], us1)) us1 <- us1[-which(us1 == dro.dat$stateb[i])]
  us2 <- unique(c(datc2$statea, datc2$stateb))
  if(is.element(dro.dat$stateb[i], us2)) us2 <- us2[-which(us2 == dro.dat$stateb[i])]
  if(is.element(dro.dat$statea[i], us2)) us2 <- us2[-which(us2 == dro.dat$statea[i])]
  if(min(c(length(us1),length(us2)))>0){
    tchange[i] <- length(intersect(us1,us2))
    tschange[i] <- length(us1)+length(us2)
  }
```

```
}

dro.dat <- cbind(dro.dat, tchange, tschange)
save(dro.dat, "dro_dat_triadicChange.RData")
```

2.4 The Bootstrapped Temporal Exponential Random Graph Model (BTERGM)

The Bootstrapped Temportal Exponential Random Graph Model (BTERGM) was initially introduced by Desmarais and Cranmer (2010) as a method of efficiently estimating network effect size and uncertainty using bootstrapped maximum pseudolikelihood esitmation (MPLE). MPLE increases the computational efficiency in estimating logitudinal network models at the cost of downwardly biased measures of uncertainty. Desmarais and Cranmer (2012) find that when taking bootstrap samples of effect estimates, consistent confidence intervals can be computed. The BTERGM offers the perfect model to test a network theory of the democratic peace for a few reasons. First, computing network effects for a dynamic network observed over 185 different time periods poses signficant computational problems. Estimating these effects through MPLE can accelerate this process significantly. Second, using an inferential approach to network analysis allows scholars to meaningfully test theories of network topology in ways that conventional generalized linear models with network covariates may be unable to. In particular, it allows for more accurate estimation of confidence intervals for the network covariates.

This model has been used for many applications in the domain of international relations, including the study of conflict (Cranmer and Desmarais 2011), sanctions (Cranmer, Heinrich and Desmarais 2014), alliances (Cranmer, Desmarais and Menninga 2012), and state prestige and status (Duque 2015). Given its state-of-the-art status and applicability to testing our theory, we believe this model is ideal. The preceeding code develops the network measures and data that will be fed to the following chunk of code, which estimates the TERGM using bootstrapped MPLE:

```
# Assuming the data is labled "dat", and the formula for the model is "formula",
  # the data files created above will be run through the following:

# For GW Specification:
  # import gw.dat object
load("gw_dat_triadicChange.RData")

  # specify gw replication equation

gw_formula <- as.formula(mzmid1_orig ~ smldem + dyaddiff + propdem +
                           dyadsyst + lndstab + contiguity + allies +
                           caprat + majpow + py + '_spline1' + '_spline2' + '_spline3')

gw_formula_network <-  as.formula(mzmid1_orig ~ smldem + dyaddiff + propdem +
                                    dyadsyst + lndstab + contiguity + allies +
                                    caprat + majpow + py + '_spline1' + '_spline2' + '_spline3
                                    tchange + tschange)
```

```r
# baseline
baseline_gw <- glm(gw_formula,
                   dat = gw.dat,
                   family = binomial)

# network model
m <- 1000
set.seed(10)
uyr <- unique(gw.dat$year)
coef.triad <- NULL
coef.triad.l <- NULL
for(i in 1:1000){
  yrsi <- sample(uyr, length(uyr), rep=TRUE)
  yrsil <- as.list(yrsi)
  inds <- unlist(lapply(yrsil, whichab, y= dat$year))
  datai <- dat[inds,]
  mod.triad <- glm(gw_formula_network, data = datai, family = binomial)
  coef.triad <- rbind(coef.triad, rbind(mod.triad$coefficients))
}

dput(list(coef = coef.triad), "GW_TERGM.txt")
save.image(file = "GW_TERGM.Rdata")

rm(list = ls())

# For DOR specification
  # import dro.dat data object
load("dro_dat_triadicChange.RData")

dro_vars <- c("statea", "stateb", "dyadid", "year", "mzmid1_orig", "jtdem",
              "propdem", "jtdemxpropdem", "jtaut", "jtautxpropdem", "mixed",
              "mixedxpropdem", "contiguity", "lndstab", "pwin_lrg", "lrgcap",
              "allies", "lnNt", "py", "_spline1", "_spline2", "_spline3")

dor_formula <- as.formula(mzmid1_orig ~ jtdem + propdem + jtdemxpropdem + jtaut +
                            jtautxpropdem + mixed + mixedxpropdem + contiguity + lndstab +
                            pwin_lrg + lrgcap + allies + lnNt + py +
                            '_spline1' + '_spline2' + '_spline3')

dor_formula_network <- as.formula(mzmid1_orig ~ jtdem + propdem + jtdemxpropdem + jtaut +
                                    jtautxpropdem + mixed + mixedxpropdem + contiguity +
                                    lndstab + pwin_lrg + lrgcap + allies + lnNt + py +
                                    '_spline1' + '_spline2' + '_spline3' +
                                    tchange + tschange)
```

```
# baseline
baseline_dor <- glm(dor_formula,
                    dat = dro.dat,
                    family = binomial)

# network model
m <- 1000
set.seed(10)
uyr <- unique(dro.dat$year)
coef.triad <- NULL
coef.triad.l <- NULL
for(i in 1:1000){
  yrsi <- sample(uyr, length(uyr), rep=TRUE)
  yrsil <- as.list(yrsi)
  inds <- unlist(lapply(yrsil, whichab, y= dat$year))
  datai <- dat[inds,]
  mod.triad <- glm(dor_formula_network, data = datai, family = binomial)
  coef.triad <- rbind(coef.triad, rbind(mod.triad$coefficients))
}

dput(list(coef = coef.triad), "DOR_TERGM.txt")
save.image(file = "DOR_TERGM.Rdata")
```

2.5 Interpreting Results

In interpreting the results from the previously described models, we will determine that our theory has empirical support if the two-stars term is positive statistically significant at the conventional $p = 0.05$ level while the triadic-closure term is negative and statistically significant at the conventional $p = 0.05$ level. However, the best test for a theory is not necessarily the traditional metric, and we intend to perform a more rigorous analysis of our model results using predictive methods (Ward, Greenhill and Bakke 2010). In this case, we intend to compare the Precision-Recall (PR) Cruves for fully specified models and models excluding the network covariates. In this case, PR Curves estimated using cross-validative techniques provide an opportunity to test model fit under the conditions of rare events and the potential for model overfitting. A significant increase in the area under the curve for the fully specified model relative to the restricted model would be evidence of improved model performance and the importance of the network effects.

3 Auxilliary Analyses

This section reviews some auxilliary analyses we will conduct, and in particular, certain robustness checks that would demonstrate added support for our theory. These auxilliary analyses range from estimating the previously discussed TERGM with distinct measures of democracy, to estimating models for fatal militarized interstate disputes (level 4 MIDs or higher), examining the dyads that

contribute most to the democratic peace using case studies, fitting the model on all MIDs as opposed to just originators, and examining the historical contexts that may drive democratic peace findings. Each of these auxilliary analyses are intended to demonstrate support for our network theory of the democratic peace. We will discuss each of these in turn.

3.1 Distinct Measures of Democracy

In an effort to determine how sensitive model results are to different measurements of joint democracy, we plan to estimate the previously described TERGM for the Dafoe, Oneal and Russett (2013) model using the following measures for the joint democracy, autocracy, and mixed-pair indicators. It is worth noting that we cannot use some measure of difference between regime scores, as the two most dmeocratic pairs (10 and 10) would differnece out to zero, and as such, we would not precisely be measuring the variable of interest. As such, the measures that we may use for this model are constrained to changing the threshold for joint democracy indicator model in such a way as to ensure that we would be more likely to detect a significant effect for the democratic peace (i.e., creating an easier test for the democratic peace):

- Joint Democracy: Dafoe, Oneal and Russett (2013) code a joint democracy as a dyad where the lowest polity value in the dyad is greater than 6. This is fairly generous, as certain states that may not be considered democracies are included as democracies. This should constitute a harder test than a coding where this threshold is higher, and as such, states may be more democratic. There is reason to think the democratic peace may be the strongest for the most democratic states based upon the normative and structural explanations provided by Maoz and Russett (1993). So, as an easier test for the democratic peace, we move this threshold higher, to 10. This is the highest polity score a state may have and would reflect a fully democratic dyad

- Joint Autocracy: This is to remain the same coding, an indicator that takes on a value of 1 if the largest polity value in the dyad is less than or equal to 6. This variable should have no bearing on the support for the democratic peace.

- Mixed Pair: This is to remain the same coding, an indicator that takes on a value of 1 if the largest poltiy value has a value less than or equal to one and the state with the highest polity value has a polity value greater than 6. This variable should have no bearing on the support for the democratic peace.

- Proportion of Democracy: If the additional support for the democratic peace found by Dafoe, Oneal and Russett (2013) is conditional on there being a higher proportion of democracies in the system, an easier test would be to increase the coding threshold for this variable. As it stands, the proporiton of democracies is a system year variable that is equal to the quotient of the number of states with polity scores of 6 or greater and the total number of states in the system. Let's assume that as the proportion of pure democracies in a system increases, keeping with Dafoe, Oneal and Russett (2013), the conditional effect of joint democracy on the outbreak of war should decrease even further. Systemically, this may reflect some norm against conflict for the "club of democracies", and as the club gets larger, the costs for breaking the norm increase. As such, we create a new coding of the proportion of democracies equal to the quotient of the number of states with polity scores of 10 or greater and the total number of states in the system.

The following chunk of code walks through the code necessary to develop these new measures and then estimate TERGMs using them.

```
library(foreign)
dat <- read.dta("dor_18162001_08252016.dta")

# Dyadic difference variable
dat$dyaddiff <- dat$lrgdem - dat$smldem
dat$dyadsyst <- dat$dyaddiff*dat$propdem
dat$caprat <- with(dat, smlcap/(smlcap + lrgcap))

# New Joint Democracy
dat$jtdem_10 <- ifelse(dat$smldem == 10, 1, 0)

# New Proportion of democracies variable
dat$perfect_dem_a <- ifelse(dat$polity2a == 10, 1, 0)
dat$perfect_dem_b <- ifelse(dat$polity2b == 10, 1, 0)
  # So, I need to loop through all of the unique states for each year and see

library(data.table)
library(dplyr)

dat <- as.data.table(dat)

new_prop_dem <- list()
for(i in 1:length(unique(dat$year))){
  year_i <- unique(dat$year)[[i]]
  dat_i <- subset(dat, year == year_i)
  dat_j <- distinct(dat_i, statea)
  dt_j <- data.table(state = dat_j$statea,
                     dem = dat_j$perfect_dem_a)
  dat_k <- distinct(dat_i, stateb)
  dt_k <- data.table(state = dat_k$stateb,
                       dem = dat_k$perfect_dem_b)
  dt <- rbind(dt_j, dt_k)
  dt <- distinct(dt, state)

  prop_dem_dt <- data.table(year = year_i,
                            prop_dem_10 = mean(dt$dem, na.rm = TRUE))

  new_prop_dem[[i]] <- prop_dem_dt
}

dat_propdem <- do.call("rbind", new_prop_dem)
```

```
dat_new <- merge(dat, dat_propdem, by = "year", all.x = TRUE)
dat <- dat_new

# create new interactions
jtdem10xpropdem10 <- with(dat, jtdem_10*prop_dem_10)
jtautxpropdem10 <- with(dat, jtaut*prop_dem_10)
mixedxpropdem10 <- with(dat, mixed*prop_dem_10)


dor_formula <- as.formula(mzmid1_orig ~ jtdem_10 + prop_dem_10 + jtdem10xpropdem10 + jtaut +
                            jtautxpropdem10 + mixed + mixedxpropdem10 + contiguity + lndstab +
                            pwin_lrg + lrgcap + allies + lnNt + py +
                            '_spline1' + '_spline2' + '_spline3')

dor_formula_network <- as.formula(mzmid1_orig ~ jtdem_10 + prop_dem_10 + jtdem10xpropdem10
                                  + jtaut + jtautxpropdem10 + mixed + mixedxpropdem10 +
                                  contiguity + lndstab + pwin_lrg + lrgcap + allies + lnNt
                                  + py + '_spline1' + '_spline2' + '_spline3' +
                                  tchange + tschange)


# baseline
baseline_dor <- glm(dor_formula,
                    dat = dro.dat,
                    family = binomial)

# network model
m <- 1000
set.seed(10)
uyr <- unique(dro.dat$year)
coef.triad <- NULL
coef.triad.l <- NULL
for(i in 1:1000){
  yrsi <- sample(uyr, length(uyr), rep=TRUE)
  yrsil <- as.list(yrsi)
  inds <- unlist(lapply(yrsil, whichab, y= dat$year))
  datai <- dat[inds,]
  mod.triad <- glm(dor_formula_network, data = datai, family = binomial)
  coef.triad <- rbind(coef.triad, rbind(mod.triad$coefficients))
}

dput(list(coef = coef.triad), "DOR_TERGM_DemCoding.txt")
save.image(file = "DOR_TERGM_DemCoding.Rdata")
```

However, it might be argued that the theoretical model provided by Gartzke and Weisiger (2013)

and empirical model provided by Dafoe, Oneal and Russett (2013) do not reflect the strongest test of the democratic peace. As such, we plan to estimate a series of TERGMs with the previously described controls and network covariates, while varying the measurement for the democratic peace. Overall, we plan to estimate 5 TERGMs with the following democracy indicators:

- Weak-Link Regime Score: The polity value equal to the lowest polity value for a dyad.
- Joint Democracy Indicator (6-threshold): This variable takes on a value of one if the lowest polity value for a dyad is equal to or greater than 6.
- Joint Democracy Indicator (8-threshold): This variable takes on a value of one if the lowest polity value for a dyad is equal to or greater than 8.
- Joint Pure Democracy Indicator (10-threshold): This variable takes on a value of one if the lowest polity value for a dyad is equal to 10, which would be evidence for a perfectly democratic dyad.
- Joint Democracy Indicator (Boix, Miller and Rosato (2012) Coding): This variable takes on a value of one if each state satisfies conditions for both contestation and participation of elections, or, recieve a value of one in the Boix, Miller and Rosato (2012) democracy data. To our knowledge this data has not been used in the democratic peace, but given its theoretically motivated coding to capture leader vulnerability and structural accountability, it seems a good fit.

The code necessary to create the above measures, tested on originators of MIDs, is as follows:

```
load("dro_dat_triadicChange.RData")

dat$jtdem_8 <- ifelse(dat$smdem > 7, 1, 0)

dat$jtdem_10 <- ifelse(dat$smldem == 10, 1, 0)

btr_dem <- read.csv("democracy.csv")

btr_dem$statea <- btr_dem$ccode
btr_dem$btr_state_a <- btr_dem$democracy
head(btr_dem)

dat_a <- merge(dat, btr_dat[,c(11, 4, 12)], by = ("statea", "year"),
               all.x = TRUE)

btr_dem$stateb <- btr_dem$ccode
btr_dem$btr_state_b <- btr_dem$democracy
head(btr_dem)

dat_b <- merge(dat_a, btr_dat[,c(13, 4, 14)], by = ("stateb", "year"),
               all.x = TRUE)
```

```
dat <- dat_b

dat$btr_joint_dem <- ifelse(dat$btr_state_a == 1 & dat$btr_state_b == 1,
                            1, 0)

save(dat, file = "dor_dat_DemMeasures.RData")


# measure 1
rm(list = ls())
load("dor_dat_DemMeasures.RData")

formula_network1 <- as.formula(mzmid1_orig ~ smldem + contiguity +
                                  lndstab + pwin_lrg + lrgcap +
                                  allies + lnNt + py +
                                  '_spline1' + '_spline2' + '_spline3' +
                                  tchange + tschange)

m <- 1000
set.seed(10)
uyr <- unique(dat$year)
coef.triad <- NULL
coef.triad.l <- NULL
for(i in 1:1000){
  yrsi <- sample(uyr, length(uyr), rep=TRUE)
  yrsil <- as.list(yrsi)
  inds <- unlist(lapply(yrsil, whichab, y= dat$year))
  datai <- dat[inds,]
  mod.triad <- glm(formula_network1,
                   data = datai,
                   family = binomial)
  coef.triad <- rbind(coef.triad, rbind(mod.triad$coefficients))
}

dput(list(coef = coef.triad), "DOR_TERGM_DemMeas1.txt")
save.image(file = "DOR_TERGM_DemMeas1.Rdata")

# measure 2
rm(list = ls())
load("dor_dat_DemMeasures.RData")

formula_network2 <- as.formula(mzmid1_orig ~ jtdem + contiguity
                                + lndstab + pwin_lrg +
                               lrgcap + allies + lnNt + py +
```

```
                                '_spline1' + '_spline2' + '_spline3' +
                                tchange + tschange)

m <- 1000
set.seed(10)
uyr <- unique(dat$year)
coef.triad <- NULL
coef.triad.l <- NULL
for(i in 1:1000){
  yrsi <- sample(uyr, length(uyr), rep=TRUE)
  yrsil <- as.list(yrsi)
  inds <- unlist(lapply(yrsil, whichab, y= dat$year))
  datai <- dat[inds,]
  mod.triad <- glm(formula_network2,
                   data = datai,
                   family = binomial)
  coef.triad <- rbind(coef.triad, rbind(mod.triad$coefficients))
}

dput(list(coef = coef.triad), "DOR_TERGM_DemMeas2.txt")
save.image(file = "DOR_TERGM_DemMeas2.Rdata")

# Measure 3
rm(list = ls())
load("dor_dat_DemMeasures.RData")

formula_network3 <- as.formula(mzmid1_orig ~ jtdem_8
                                + contiguity + lndstab + pwin_lrg +
                                lrgcap + allies + lnNt + py +
                                '_spline1' + '_spline2' + '_spline3' +
                                tchange + tschange)


m <- 1000
set.seed(10)
uyr <- unique(dat$year)
coef.triad <- NULL
coef.triad.l <- NULL
for(i in 1:1000){
  yrsi <- sample(uyr, length(uyr), rep=TRUE)
  yrsil <- as.list(yrsi)
  inds <- unlist(lapply(yrsil, whichab, y= dat$year))
  datai <- dat[inds,]
  mod.triad <- glm(formula_network3,
                   data = datai,
```

```
                    family = binomial)
  coef.triad <- rbind(coef.triad, rbind(mod.triad$coefficients))
}

dput(list(coef = coef.triad), "DOR_TERGM_DemMeas3.txt")
save.image(file = "DOR_TERGM_DemMeas3.Rdata")

# measure 4
rm(list = ls())
load("dor_dat_DemMeasures.RData")

formula_network4 <- as.formula(mzmid1_orig ~ jtdem_10 + contiguity
                               + lndstab + pwin_lrg +
                               lrgcap + allies + lnNt + py +
                               '_spline1' + '_spline2' + '_spline3' +
                               tchange + tschange)

m <- 1000
set.seed(10)
uyr <- unique(dat$year)
coef.triad <- NULL
coef.triad.l <- NULL
for(i in 1:1000){
  yrsi <- sample(uyr, length(uyr), rep=TRUE)
  yrsil <- as.list(yrsi)
  inds <- unlist(lapply(yrsil, whichab, y= dat$year))
  datai <- dat[inds,]
  mod.triad <- glm(formula_network4,
                   data = datai,
                   family = binomial)
  coef.triad <- rbind(coef.triad, rbind(mod.triad$coefficients))
}

dput(list(coef = coef.triad), "DOR_TERGM_DemMeas4.txt")
save.image(file = "DOR_TERGM_DemMeas4.Rdata")

# measure 5
rm(list = ls())
load("dor_dat_DemMeasures.RData")

formula_network5 <- as.formula(mzmid1_orig ~ btr_joint_dem + contiguity
                               + lndstab + pwin_lrg +
                               lrgcap + allies + lnNt + py +
                               '_spline1' + '_spline2' + '_spline3' +
                               tchange + tschange)
```

```
m <- 1000
set.seed(10)
uyr <- unique(dat$year)
coef.triad <- NULL
coef.triad.l <- NULL
for(i in 1:1000){
  yrsi <- sample(uyr, length(uyr), rep=TRUE)
  yrsil <- as.list(yrsi)
  inds <- unlist(lapply(yrsil, whichab, y= dat$year))
  datai <- dat[inds,]
  mod.triad <- glm(formula_network5,
                   data = datai,
                   family = binomial)
  coef.triad <- rbind(coef.triad, rbind(mod.triad$coefficients))
}

dput(list(coef = coef.triad), "DOR_TERGM_DemMeas5.txt")
save.image(file = "DOR_TERGM_DemMeas5.Rdata")
```

3.2 Model Estimating using Fatal MIDs

A commonly used robustness check for studies on MIDs is to change the coding of the dependent variable from any level MID, 1 or greater, to a "fatal" MID which takes on a value of 4 of greater. In an effort to demonstrate additional robustness for our primary analyses, we intend to estimate our primary TERGM specifications using fatal MIDs as the outcome and the network used to generate our network statistics. These network statistics are generated as follows:

```
# Load relevant packages
library(foreign)
library(statnet)
library(data.table)
library(DataCombine)
# Load data from Dafoe, Oneal, and Russett
  # This data was prepared according to their data cleaning script
  # and then exported from Stata12
dat <- read.dta("dor_18162001_08252016.dta")

# Dyadic difference variable
dat$dyaddiff <- dat$lrgdem - dat$smldem
dat$dyadsyst <- dat$dyaddiff*dat$propdem
dat$caprat <- with(dat, smlcap/(smlcap + lrgcap))
dat$majpow <- ifelse(dat$majpow1 | dat$majpow2 == 1, 1, 0)
```

```
for(i in 1:nrow(dat)){
  dat1 <- subset(dat, dat$year==dat$year[i])
  dat1mid <- subset(dat1, dat1$mzfatal1_orig==1)
  datc1 <- subset(dat1mid, as.numeric(dat1mid$statea == dat$statea[i]) +
                    as.numeric(dat1mid$stateb == dat$statea[i]) > 0)
  datc2 <- subset(dat1mid, as.numeric(dat1mid$statea == dat$stateb[i]) +
                    as.numeric(dat1mid$stateb == dat$stateb[i]) > 0)
  us1 <- unique(c(datc1$statea, datc1$stateb))
  if(is.element(dat$statea[i], us1)) us1 <- us1[-which(us1 == dat$statea[i])]
  if(is.element(dat$stateb[i], us1)) us1 <- us1[-which(us1 == dat$stateb[i])]
  us2 <- unique(c(datc2$statea, datc2$stateb))
  if(is.element(dat$stateb[i], us2)) us2 <- us2[-which(us2 == dat$stateb[i])]
  if(is.element(dat$statea[i], us2)) us2 <- us2[-which(us2 == dat$statea[i])]
  if(min(c(length(us1),length(us2)))>0){
    tchange[i] <- length(intersect(us1,us2))
    tschange[i] <- length(us1)+length(us2)
  }
}



dor_formula <- as.formula(mzfatal1_orig ~ jtdem + propdem + jtdemxpropdem +
                            jtaut + jtautxpropdem + mixed + mixedxpropdem +
                            contiguity + lndstab +
                            pwin_lrg + lrgcap + allies + lnNt + py +
                            '_spline1' + '_spline2' + '_spline3')

dor_formula_network <- as.formula(mzfatal1_orig ~ jtdem + propdem +
                                    jtdemxpropdem + jtaut + jtautxpropdem +
                                    mixed + mixedxpropdem + contiguity +
                                    lndstab + pwin_lrg + lrgcap + allies +
                                    lnNt + py + '_spline1' + '_spline2' +
                                    '_spline3' + tchange + tschange)


# baseline
baseline_dor <- glm(dor_formula,
                    dat = dro.dat,
                    family = binomial)

# network model
m <- 1000
set.seed(10)
uyr <- unique(dro.dat$year)
coef.triad <- NULL
```

```
coef.triad.l <- NULL
for(i in 1:1000){
  yrsi <- sample(uyr, length(uyr), rep=TRUE)
  yrsil <- as.list(yrsi)
  inds <- unlist(lapply(yrsil, whichab, y= dat$year))
  datai <- dat[inds,]
  mod.triad <- glm(dor_formula_network, data = datai, family = binomial)
  coef.triad <- rbind(coef.triad, rbind(mod.triad$coefficients))
}

dput(list(coef = coef.triad), "DOR_TERGM_FatalOrig.txt")
save.image(file = "DOR_TERGM_FatalOrig.Rdata")
rm(list = ls())

# All fatal mids
dat <- read.dta("dor_18162001_08252016.dta")

# Dyadic difference variable
dat$dyaddiff <- dat$lrgdem - dat$smldem
dat$dyadsyst <- dat$dyaddiff*dat$propdem
dat$caprat <- with(dat, smlcap/(smlcap + lrgcap))
dat$majpow <- ifelse(dat$majpow1 | dat$majpow2 == 1, 1, 0)


for(i in 1:nrow(dat)){
  dat1 <- subset(dat, dat$year==dat$year[i])
  dat1mid <- subset(dat1, dat1$mzfatal1_all==1)
  datc1 <- subset(dat1mid, as.numeric(dat1mid$statea == dat$statea[i]) +
                    as.numeric(dat1mid$stateb == dat$statea[i]) > 0)
  datc2 <- subset(dat1mid, as.numeric(dat1mid$statea == dat$stateb[i]) +
                    as.numeric(dat1mid$stateb == dat$stateb[i]) > 0)
  us1 <- unique(c(datc1$statea, datc1$stateb))
  if(is.element(dat$statea[i], us1)) us1 <- us1[-which(us1 == dat$statea[i])]
  if(is.element(dat$stateb[i], us1)) us1 <- us1[-which(us1 == dat$stateb[i])]
  us2 <- unique(c(datc2$statea, datc2$stateb))
  if(is.element(dat$stateb[i], us2)) us2 <- us2[-which(us2 == dat$stateb[i])]
  if(is.element(dat$statea[i], us2)) us2 <- us2[-which(us2 == dat$statea[i])]
  if(min(c(length(us1),length(us2)))>0){
    tchange[i] <- length(intersect(us1,us2))
    tschange[i] <- length(us1)+length(us2)
  }
}

dor_formula <- as.formula(mzfatal1_all ~ jtdem + propdem + jtdemxpropdem +
                            jtaut + jtautxpropdem + mixed + mixedxpropdem +
```

```
                              contiguity + lndstab +
                              pwin_lrg + lrgcap + allies + lnNt + py +
                              '_spline1' + '_spline2' + '_spline3')

dor_formula_network <- as.formula(mzfatal1_all ~ jtdem + propdem + jtdemxpropdem
                                    + jtaut + jtautxpropdem + mixed + mixedxpropdem
                                    + contiguity + lndstab + pwin_lrg + lrgcap
                                    + allies + lnNt + py + '_spline1' + '_spline2'
                                    + '_spline3' + tchange + tschange)

# baseline
baseline_dor <- glm(dor_formula,
                    dat = dro.dat,
                    family = binomial)

# network model
m <- 1000
set.seed(10)
uyr <- unique(dro.dat$year)
coef.triad <- NULL
coef.triad.l <- NULL
for(i in 1:1000){
  yrsi <- sample(uyr, length(uyr), rep=TRUE)
  yrsil <- as.list(yrsi)
  inds <- unlist(lapply(yrsil, whichab, y= dat$year))
  datai <- dat[inds,]
  mod.triad <- glm(dor_formula_network, data = datai, family = binomial)
  coef.triad <- rbind(coef.triad, rbind(mod.triad$coefficients))
}

dput(list(coef = coef.triad), "DOR_TERGM_FatalAll.txt")
save.image(file = "DOR_TERGM_FatalAll.Rdata")
```

3.3 Examining the Dyads that Contribute Most to the Democratic Peace

In an effort to find the "smoking gun" for our theory and the corresponding results, we will examine the dyads that contribute most to the democratic peace and provide a closer examination of them using case studies. In particular, when looking at the dyads that contribute most to the democratic peace, we will pay attention as to whether the two dyads did not fight because they were democratic, or because of their particular location in the conflict network. Did these dyads have common enemies? Did they have certain positions that prohibited them from fighting? Or, did regime type truely determine their pacificity? To answer these questions, and to find these cases, we will iteratively remove each dyad-year from the data and then fit a fully specified logistic regression on

each reduced dataset. We will then3 examine the change in the baseline joint democracy coefficient.[1] In particular, we will use the weakest-link regime score as the variable of interest. Code for this analysis is as follows:

```
load("dro_dat_triadicChange.RData")
  # Assuming name is dat


formula <- as.formula(mzmid1_orig ~ smldem + contiguity + lndstab + pwin_lrg +
                        lrgcap + allies + lnNt + py +
                        '_spline1' + '_spline2' + '_spline3')

# Will consider parallelizing based upon computational time

coef_change <- list()
se_change <- list()
BIC_change <- list()

for(i in 1:nrow(dat)){
  dat_temp <- dat[-c(i),]
  mod <- glm(formula,
             family = binomial,
             data = dat_temp)
  coef <- summary(mod)$coefficients[2,1]
  se <- summary(mod)$coefficients[2,2]
  BIC <- BIC(mod)
  coef_change[[i]] <- coef
  se_change[[i]] <- se
  BIC_change[[i]] <- BIC
}

dat$coef_change <- unlist(coef_change)
dat$se_change <- unlist(se_change)
dat$BIC_change <- unlist(BIC_change)

mod <- glm(formula,
           family = binomial,
           data = dat)

dat$coef <- summary(mod)$coefficients[2,1]
dat$se <- summary(mod)$coefficients[2,2]
dat$BIC <- BIC(mod)

dat$importance_coef <- with(dat, coef - coef_change)
```

[1] Given that we are estimating these models without the network statistics, we need not use the TERGM.

```
dat$importance_se <- with(dat, se - se_change)
dat$importance_BIC <- with(dat, BIC - BIC_change)

# Largest negative values for the difference in coefficient from baseline
  # is what we want
five_important_dyads <- head(dat[order(-dat$importance_coef),])
```

3.4 Examining Temporal Heterogeneity in the Democratic Peace

Scholars are increasingly concerned with the temporal stability of effects estimated on time-series cross-sectional data (Wawro and Katznelson 2014), and in particular, the temporal stability of many models of interstate conflict (Jenke and Gelpi 2016). As such, an additional approach to uncover evidence for a network-based theory of the democratic peace is to explore the temporal heterogeneity associated with the effect of joint democracy on MID-propensity. One might expect certain historical contexts to drive the democratic peace during certain temporal windows. For example, during the Cold War, there may be resounding evidence for the democratic peace. However, is such evidence a function of a dyadic-attribute such as joint regime type, or some larger systemic network context, such as common enemies.

To explore such possibilities, we will estimate a Structured Additive Regression (STAR) Model. These models can be understood as an extension of General Additive Models (GAMs), with the advantage of allowing analysts to incorporate nonparametric functions of covariates when the functional form of a variable's relationship with the outcome is unknown. STAR is implemented through `BayesX` software, with an `R` interface, and uses MCMC-based estimation (Brezger, Kneib and Lang 2003).

This model specification will mirror that used to generate the "important dyads" previously described, while specifying that the democracy variable is a function of the year. Code for this exercise is approximated below:

```
load("dro_dat_triadicChange.RData")

library(BayesXsrc)
library(R2BayesX)


formula <- as.formula(mzmid1_orig ~ sx(year, by = smldem, center = TRUE) +
                        contiguity + lndstab + pwin_lrg +
                        lrgcap + allies + lnNt + py +
                        '_spline1' + '_spline2' + '_spline3')

mod <- bayesx(formula,
              method = "MCMC",
              iterations = 25000L,
              burnin = 5000L,
              seed = 1234,
```

```
              CI = "MCMCbootstrap",
              data = dat,
              family = binomial)

plot(mod, term = "sx(year):DEMOCRACY",
     xlab = "Year",
     ylab = "Coefficient of Weak-Link Democracy Score")
```

Once we have extracted the time-series of democracy coefficients, we will perform a historical and contextual analysis of periodicity in the effect size and direction. This is to say, we will examine the historical context that exists between change-points in the time-series of the coefficient and look at the events that lead to significant changes in the coefficient. In particular, we will look at the periods where this coefficient is at its most-negative and determine whether joint democracy was driving the result, or if historically contingent network features were at play.

3.5 Examining All Participants

The previous analyses we have described have examined the network of MID initiators, as suggested by Bennett and Stam (2004), who argue that looking at all MIDs, including joiners, changes the sitution for countries who might join or be forced to join. However, thinking of this as a network of MID initiators may leave a fairly sparse network that does not include the network-based dependencies that lead states to join MIDs. As such, we perform additional analyses based upon the primary analyses that looks at all participants. We may expect network effects to be more present when examining all participants, including "joiners". To perform this analysis, we look at all MIDs and fatal MIDs using our replication of the Dafoe, Oneal and Russett (2013) specification. The code for this analysis is as folows:

```
# Set local working directory
setwd("~/Box Sync/NetworkDemocraticPeace")

# Load relevant packages
library(foreign)
library(statnet)
library(data.table)
library(DataCombine)
# Load data from Dafoe, Oneal, and Russett
  # This data was prepared according to their data cleaning script
  # and then exported from Stata12
dat <- read.dta("dor_18162001_08252016.dta")

# Dyadic difference variable
dat$dyaddiff <- dat$lrgdem - dat$smldem
dat$dyadsyst <- dat$dyaddiff*dat$propdem
dat$caprat <- with(dat, smlcap/(smlcap + lrgcap))
dat$majpow <- ifelse(dat$majpow1 | dat$majpow2 == 1, 1, 0)
```

```r
dro_vars <- c("statea", "stateb", "dyadid", "year", "mzmid1_all", "jtdem",
              "propdem", "jtdemxpropdem", "jtaut", "jtautxpropdem", "mixed",
              "mixedxpropdem", "contiguity", "lndstab", "pwin_lrg", "lrgcap",
              "allies", "lnNt", "py", "_spline1", "_spline2", "_spline3")

dro.dat <- NULL
for(i in dro_vars){
  dro.dat <- cbind(dro.dat, dat[,which(names(dat)==i)])
}

dro.dat <- as.data.frame(dro.dat)
names(dro.dat) <- dro_vars
dro.dat <- na.omit(dro.dat)

# Create triadic change stats
tchange <- numeric(nrow(dro.dat))
tschange <- numeric(nrow(dro.dat))

for(i in 1:nrow(dro.dat)){
  dat1 <- subset(dro.dat, dro.dat$year==dro.dat$year[i])
  dat1mid <- subset(dat1, dat1$mzmid1_all==1)
  datc1 <- subset(dat1mid, as.numeric(dat1mid$statea == dro.dat$statea[i]) +
                    as.numeric(dat1mid$stateb == dro.dat$statea[i]) > 0)
  datc2 <- subset(dat1mid, as.numeric(dat1mid$statea == dro.dat$stateb[i]) +
                    as.numeric(dat1mid$stateb == dro.dat$stateb[i]) > 0)
  us1 <- unique(c(datc1$statea, datc1$stateb))
  if(is.element(dro.dat$statea[i], us1)) us1 <- us1[-which(us1 == dro.dat$statea[i])]
  if(is.element(dro.dat$stateb[i], us1)) us1 <- us1[-which(us1 == dro.dat$stateb[i])]
  us2 <- unique(c(datc2$statea, datc2$stateb))
  if(is.element(dro.dat$stateb[i], us2)) us2 <- us2[-which(us2 == dro.dat$stateb[i])]
  if(is.element(dro.dat$statea[i], us2)) us2 <- us2[-which(us2 == dro.dat$statea[i])]
  if(min(c(length(us1),length(us2)))>0){
    tchange[i] <- length(intersect(us1,us2))
    tschange[i] <- length(us1)+length(us2)
  }
}

dro.dat <- cbind(dro.dat, tchange, tschange)
save(dro.dat, "dro_dat_allMids.RData")

dor_formula <- as.formula(mzmid1_all ~ jtdem + propdem + jtdemxpropdem +
                            jtaut + jtautxpropdem + mixed +
                            mixedxpropdem + contiguity + lndstab +
                            pwin_lrg + lrgcap + allies + lnNt + py +
                            '_spline1' + '_spline2' + '_spline3')
```

```
dor_formula_network <- as.formula(mzmid1_all ~ jtdem + propdem + jtdemxpropdem +
                                    jtaut + jtautxpropdem + mixed +
                                    mixedxpropdem + contiguity +
                                    lndstab + pwin_lrg + lrgcap + allies + lnNt + py +
                                    '_spline1' + '_spline2' + '_spline3' +
                                    tchange + tschange)

# baseline
baseline_dor <- glm(dor_formula,
                    dat = dro.dat,
                    family = binomial)

# network model
m <- 1000
set.seed(10)
uyr <- unique(dro.dat$year)
coef.triad <- NULL
coef.triad.l <- NULL
for(i in 1:1000){
  yrsi <- sample(uyr, length(uyr), rep=TRUE)
  yrsil <- as.list(yrsi)
  inds <- unlist(lapply(yrsil, whichab, y= dat$year))
  datai <- dat[inds,]
  mod.triad <- glm(dor_formula_network, data = datai, family = binomial)
  coef.triad <- rbind(coef.triad, rbind(mod.triad$coefficients))
}

dput(list(coef = coef.triad), "DOR_TERGM_allMids.txt")
save.image(file = "DOR_TERGM_allMids.Rdata")

# Now for all fatal mids
rm(list = ls())


# Load data from Dafoe, Oneal, and Russett
  # This data was prepared according to their data cleaning script
  # and then exported from Stata12
dat <- read.dta("dor_18162001_08252016.dta")

# Dyadic difference variable
dat$dyaddiff <- dat$lrgdem - dat$smldem
dat$dyadsyst <- dat$dyaddiff*dat$propdem
dat$caprat <- with(dat, smlcap/(smlcap + lrgcap))
dat$majpow <- ifelse(dat$majpow1 | dat$majpow2 == 1, 1, 0)
```

```
dro_vars <- c("statea", "stateb", "dyadid", "year", "mzfatal1_all", "jtdem",
              "propdem", "jtdemxpropdem", "jtaut", "jtautxpropdem", "mixed",
              "mixedxpropdem", "contiguity", "lndstab", "pwin_lrg", "lrgcap",
              "allies", "lnNt", "py", "_spline1", "_spline2", "_spline3")

dro.dat <- NULL
for(i in dro_vars){
  dro.dat <- cbind(dro.dat, dat[,which(names(dat)==i)])
}

dro.dat <- as.data.frame(dro.dat)
names(dro.dat) <- dro_vars
dro.dat <- na.omit(dro.dat)

# Create triadic change stats
tchange <- numeric(nrow(dro.dat))
tschange <- numeric(nrow(dro.dat))

for(i in 1:nrow(dro.dat)){
  dat1 <- subset(dro.dat, dro.dat$year==dro.dat$year[i])
  dat1mid <- subset(dat1, dat1$mzfatal1_all==1)
  datc1 <- subset(dat1mid, as.numeric(dat1mid$statea == dro.dat$statea[i]) +
                    as.numeric(dat1mid$stateb == dro.dat$statea[i]) > 0)
  datc2 <- subset(dat1mid, as.numeric(dat1mid$statea == dro.dat$stateb[i]) +
                    as.numeric(dat1mid$stateb == dro.dat$stateb[i]) > 0)
  us1 <- unique(c(datc1$statea, datc1$stateb))
  if(is.element(dro.dat$statea[i], us1)) us1 <- us1[-which(us1 == dro.dat$statea[i])]
  if(is.element(dro.dat$stateb[i], us1)) us1 <- us1[-which(us1 == dro.dat$stateb[i])]
  us2 <- unique(c(datc2$statea, datc2$stateb))
  if(is.element(dro.dat$stateb[i], us2)) us2 <- us2[-which(us2 == dro.dat$stateb[i])]
  if(is.element(dro.dat$statea[i], us2)) us2 <- us2[-which(us2 == dro.dat$statea[i])]
  if(min(c(length(us1),length(us2)))>0){
    tchange[i] <- length(intersect(us1,us2))
    tschange[i] <- length(us1)+length(us2)
  }
}

dro.dat <- cbind(dro.dat, tchange, tschange)
save(dro.dat, "dro_dat_allFatalMids.RData")

dor_formula <- as.formula(mzfatal1_all ~ jtdem + propdem + jtdemxpropdem + jtaut +
                            jtautxpropdem + mixed + mixedxpropdem + contiguity +
                            lndstab + pwin_lrg + lrgcap + allies + lnNt + py +
                            '_spline1' + '_spline2' + '_spline3')
```

```
dor_formula_network <- as.formula(mzfatal1_all ~ jtdem + propdem + jtdemxpropdem +
                                    jtaut +jtautxpropdem + mixed + mixedxpropdem +
                                    contiguity +  lndstab + pwin_lrg + lrgcap +
                                    allies + lnNt + py +
                                    '_spline1' + '_spline2' + '_spline3' +
                                    tchange + tschange)

# baseline
baseline_dor <- glm(dor_formula,
                    dat = dro.dat,
                    family = binomial)

# network model
m <- 1000
set.seed(10)
uyr <- unique(dro.dat$year)
coef.triad <- NULL
coef.triad.l <- NULL
for(i in 1:1000){
  yrsi <- sample(uyr, length(uyr), rep=TRUE)
  yrsil <- as.list(yrsi)
  inds <- unlist(lapply(yrsil, whichab, y= dat$year))
  datai <- dat[inds,]
  mod.triad <- glm(dor_formula_network, data = datai, family = binomial)
  coef.triad <- rbind(coef.triad, rbind(mod.triad$coefficients))
}

dput(list(coef = coef.triad), "DOR_TERGM_allFatalMids.txt")
save.image(file = "DOR_TERGM_allFatalMids.Rdata")
```

References

Beck, Nathaniel, Jonathan N Katz and Richard Tucker. 1998. "Taking time seriously: Time-series-cross-section analysis with a binary dependent variable." *American Journal of Political Science* 42(4):1260–1288.

Bennett, D Scott and Alan Stam. 2004. *The Behavioral Origins of War*. Ann Arbor: University of Michigan Press.

Boix, Carles, Michael Miller and Sebastian Rosato. 2012. "A complete data set of political regimes, 1800–2007." *Comparative Political Studies* p. 0010414012463905.

Braumoeller, Bear F and Austin Carson. 2011. "Political irrelevance, democracy, and the limits of militarized conflict." *Journal of Conflict Resolution* 55(2):292–320.

Brezger, Andreas, Thomas Kneib and Stefan Lang. 2003. "BayesX: Analysing Bayesian structured additive regression models.".

Cederman, Lars-Erik and Kristian Skrede Gleditsch. 2004. "Conquest and regime change: An evolutionary model of the spread of democracy and peace." *International Studies Quarterly* 48(3):603–629.

Cranmer, Skyler J and Bruce A Desmarais. 2011. "Inferential network analysis with exponential random graph models." *Political Analysis* 19(1):66–86.

Cranmer, Skyler J, Bruce A Desmarais and Elizabeth J Menninga. 2012. "Complex dependencies in the alliance network." *Conflict Management and Peace Science* 29(3):279–313.

Cranmer, Skyler J, Tobias Heinrich and Bruce A Desmarais. 2014. "Reciprocity and the structural determinants of the international sanctions network." *Social Networks* 36:5–22.

Dafoe, Allan, John R Oneal and Bruce Russett. 2013. "The democratic peace: Weighing the evidence and cautious inference." *International Studies Quarterly* 57(1):201–214.

Desmarais, Bruce A and Skyler J Cranmer. 2010. Consistent confidence intervals for maximum pseudolikelihood estimators. In *Proceedings of the Neural Information Processing Systems 2010 Workshop on Computational Social Science and the Wisdom of Crowds*.

Desmarais, Bruce A and Skyler J Cranmer. 2012. "Statistical mechanics of networks: Estimation and uncertainty." *Physica A: Statistical Mechanics and its Applications* 391(4):1865–1876.

Duque, Marina. 2015. "Recognizing International Status.".

Gartzke, Erik and Alex Weisiger. 2013. "Permanent friends? Dynamic difference and the democratic peace." *International Studies Quarterly* 57(1):171–185.

Hegre, Håvard. 2014. "Democracy and armed conflict." *Journal of Peace Research* p. 0022343313512852.

Jenke, Libby and Christopher Gelpi. 2016. "Theme and Variations Historical Contingencies in the Causal Model of Interstate Conflict." *Journal of Conflict Resolution* p. 0022002715615190.

Levy, Jack S. 1998. "The causes of war and the conditions of peace." *Annual Review of Political Science* 1(1):139–165.

Maoz, Zeev and Bruce Russett. 1993. "Normative and Structural Causes of Democratic Peace, 1946–1986." *American Political Science Review* 87(03):624–638.

Maoz, Zeev, Ranan D Kuperman, Lesley Terris and Ilan Talmud. 2006. "Structural equivalence and international conflict a social networks analysis." *Journal of Conflict Resolution* 50(5):664–689.

Raknerud, Arvid and Håvard Hegre. 1997. "The hazard of war: Reassessing the evidence for the democratic peace." *Journal of Peace Research* 34(4):385–404.

Ward, Michael D, Brian D Greenhill and Kristin M Bakke. 2010. "The perils of policy by p-value: Predicting civil conflicts." *Journal of Peace Research* 47(4):363–375.

Wawro, Gregory J and Ira Katznelson. 2014. "Designing historical social scientific inquiry: how parameter heterogeneity can bridge the methodological divide between quantitative and qualitative approaches." *American Journal of Political Science* 58(2):526–546.