Press Freedom Homophily in the Tie Structure of the Global Internet

Hyunjin Seo
*William Allen White School of Journalism and
Mass Communications
University of Kansas
Lawrence, KS 66045
Email: hseo@ku.edu*

Stuart Thorson
*The Maxwell School
Syracuse University (emeritus)
Syracuse, NY 13244
Email: thorson@syr.edu*

Abstract

We investigate homophily in the tie structure of the global Internet by estimating Exponential Random Graph models. Specifically, we analyze the extent to which different variables including Gross National Income, geographic proximity, political regime type, and press freedom rating account for the pattern of direct country-to-country Internet connections. Results show that for 2011-2014, but not before, press freedom homophily is significantly predictive of the presence (or absence) of country-to-country Internet connections even when controlling for geographic proximity, bandwidth, and whether or not a country is democratic. The findings provide insights into changes in press freedom around the world and the evolution of the global Internet structure.

1. Introduction

The revival of nationalist politics in Europe and the United States has been accompanied by differing claims involving the role of the press in a democracy. For example, on February 17, 2017, U.S. President Donald Trump tweeted "The FAKE NEWS media (failing @nytimes, @NBCNews, @ABC, @CBS, @CNN) is not my enemy, it is the enemy of the American People!" While politicians have often felt themselves in an adversarial relationship with the press, "the enemy of the people" phrasing struck many crossing a line in a country whose constitution prohibits government interference with a free press. The next day, German Chancellor Angela Merkel, speaking to an audience including U.S. Vice President Mike Pence, offered a contrasting position on press/government relations saying, "I stand by a free and independent press and have great respect for journalists. We've always done well in Germany when we mutually respect each other" [1].

Interestingly, these conversations about *fake news* and *press freedom* are taking place in a communication environment globalized by the Internet. Moreover, increasingly available and affordable digital communication networks enabled by the Internet allow actors other than traditional intermediaries such as governments or mass media to play an important role in producing and sharing news and information. However, different levels of access to the Internet and freedom to use the Internet as well as press freedom influence levels of citizen participation in important and relevant discourses in politics or other areas of citizen lives. In particular, country-level Internet infrastructure and policy are important factors in this. For example, North Korea's government policies deny Internet access to most of its citizens. The Chinese government generally restricts access to Google and most of its services. South Korea's national security law prohibits citizens access to North Korean websites and limits discussions of North Korea over social media. Internet access has become an integral component of freedom of press and expression even as much of the world's population lacks the resources and/or skills to easily access it.

In this sense, it is important to better understand structural changes of the global Internet and how these changes are associated with changes in other aspects of society. Against this backdrop, this paper examines several Exponential Random Graph (ERG) models to understand the extent to which social, political and economic variables including press freedom, regime type, and economic performance can account for the pattern of direct country-to-country Internet connection.[1] Our findings suggest that in the early days (2002) of the global Internet, macro economic performance, per capita domestic Internet users, and homophily with regard to regime type (whether or not a country is democratic) and geographic location are statistically meaningful predictors of structure. However, in the less centralized global Internet of 2014, regime type homophily is no longer predictive though freedom of press homophily has become significant. Our paper provides a more nuanced understanding of interplay between structural aspects of the global Internet and press freedom and regime type.

---

1. We use the existence of a direct positive bandwidth capacity link between two countries to indicate those two countries are connected even if it were to be the case that no traffic flowed over that connection in a given time period. More on this in the Method section.

HICSS

2. Background

The digital network now known as the Internet can be thought of as beginning with Kleinrock's 1961 paper [2] on packet switching. By 1969, his networking ideas had been realized in the initial ARPANET connecting machines located in Menlo Park, California (Stanford Research Institute); University of California, Los Angeles; University of California, Santa Barbara; and the University of Utah.

By 2002 the Internet was truly international, and 187 countries shared a high speed direct connection with at least one other country with a total global bandwidth capacity of 0.9 terabits per second (Tbps). By 2014 there were 201 countries connected and total global bandwidth capacity had grown to 137.3 Tbps.

The rapid growth in the size of the global Internet is not the whole story. The structure of the Internet has also undergone considerable change from being highly centralized around the United States to becoming increasingly decentralized with increasing numbers of countries having direct connections with more than one other country. In 2002, the median number of direct connections across all countries was 2 while the most direct connections any country had was 124 (U.S.). The Freeman centralization score [3] is an index of how central the most central node in a network is relative to all other nodes normalized by the theoretical maximum centralization of the network. Thus scores can range between 0 (maximally decentralized) and 1 (maximally centralized). In 2002, the score for the global Internet, using number of direct connections as the index of a country's centrality, was 0.64. Moving forward to 2014, the median number of direct connections had more than doubled to 5. Great Britain now had the largest number of direct connections (100). Importantly, the Freeman centrality score in 2014 had dropped to 0.45 indicating increased decentralization with regard to the number of direct connections held by countries connected in the global Internet.

In this paper our interest is in considering some of the variables associated with these structural changes of the global Internet. In doing so, we apply Exponential Random Graph (ERG) modeling, which has gained attention in the field of communication [4]–[6]. ERGM is used to examine interdependent mechanisms that affect how network ties are created, sustained, and dissolved, and to analyze the statistical likelihood that a particular network structure will be observed. While ERGM was initially developed in the 1990s [7], it has taken off in recent years due to excellent expository texts along with packages for use in statistical environments such as *R* [5], [8]–[11].

[6] applied ERGM to analyzing how individual and network-level factors influence the formation of online social relationships. They found that the amount of time spent online was positively associated with the likelihood of online relationship formation. Another significant predictor of the emergence of online relationship ties in their study was network pressure toward balance, "individuals tending to form relationships with others who have relationships in common" [p. 180]. [4] used ERGM to examine what factors influence structural properties of discussion networks.

Previous studies of democracy and Internet adoption examined whether forms of political institutions influence the government's Internet adoption or whether the government's Internet adoption affects its democratic process [12]–[17]. With empirical and theoretical studies in this area putting forward somewhat conflicting arguments, there are two primary schools of thoughts. The first group contends that the type of political institution is closely associated with the country's adoption of the Internet and that democratic countries are more likely to adopt the Internet [12], [14], [15]. For example, based on data from 1991 to 2001 of about 200 countries, [15] concluded that democracies adopt the Internet at a much faster pace than autocracies. Similarly, Guillen and Suarez's study [12] showed that a country's democracy score was significantly associated with the number of Internet users and hosts in that country.

In comparison, the second group suggests that the Internet adoption has little to do with advancements in democracy arguing that the association between the Internet and democratic development is overestimated [17]. Based on an empirical analysis of Internet use in authoritarian regimes from 1993 to 2010, [17] concluded that movements toward democracy were more frequent in autocratic countries with low Internet penetration. According to that study, autocracies which are concerned about the domestic information environment tend to introduce the Internet more actively than other autocracies.

While regime type has been studied to a certain extent, press freedom has rarely been examined in this context. Examining press freedom, not just regime type, is important as a country's press freedom can improve or deteriorate despite the country maintaining the same democratic status in terms of political institutions. For example, South Korea's press freedom ratings significantly plunged during Lee Myung-bak and Park Geun-hye administrations primiarily due to their attempts to censor online content. Our study aims to address this gap using the press freedom index developed by Freedom House, a U.S.-based non-governmental organization for promoting democracy and press freedom around the world. Freedom House publishes their Freedom of the Press report every year evaluating the status of press freedom in countries across the globe. The press freedom rating – *free*, *partly free*, or *not free* is based on scores each country receives in the legal, political and economic environments in which media operate [18].

The Freedom of the Press index has been widely studied in social science research [19]–[22]. For example, in their study of 115 countries over the period of 2002-2010, [19] found a positive correlation between press freedom and economic growth. These authors analyzed foreign direct investment (FDI) and gross domestic product (GDP) along with the press freedom index. Karlekar and Becker conducted a systematic investigation of the relationship between democracy and press freedom with a particular focus on the Freedom House measures. They concluded that while there has been a strong positive correlation between press

freedom and democracy measures, "Most often, changes in the state of media freedom have happened in tandem with changes in broader freedoms, therefore making it a sensitive indicator of the overall health of a democracy" [22, p. 32]. In the case of the 2014 global Internet, as discussed in the Method section, Spearman's ρ between the dichotomous regime type variable and press freedom status is .73 and generally consistent with [22].

Sobel, Dutta and Roy [23] looked at the extent to which press freedom in one country spills over to neighboring countries. Using time series analysis on a sample of 102 countries over the 1994 to 2003 period, they concluded that "a country catches approximately one-fifth of the press freedom of its neighbors" [23, p. 141]. Neighbor here refers to countries sharing a land border. The global Internet has arguably enabled countries some distance from one another to become information neighbors. If so, we might expect press freedom homophily to be associated with direct Internet ties.

3. Method

Data: Our primary analysis examines relations between the structure of Internet connections between countries in 2014[2] and economic performance (Gross National Income), number of domestic Internet users per capita, and homophily with regard to (i) geographic location, (ii) regime type (democratic or non-democratic) and (iii) the degree to which the country has a free press. We combine data from several sources. Values for country population (used in per capita calculations), gross national income, *GNI*[3], and Internet Users per 100 population measures are from the World Bank [24].[4] Regime type (*RegimeType*) is calculated from data provided by the Institut for Statskundskab [25]. Country global Internet connectivity measures are from capacity data curated by TeleGeography [26] and purchased from them. In addition, International Internet bandwidth is considered a useful and relevant indicator of global Internet connectedness, especially as previous research has shown that Internet bandwidth and Internet traffic tend to correlate [27]. Finally, the press freedom measure is from Freedom House [28].

[29]'s regime type measurements focus on electoral participation and code country government types into one of seven lexical categories: (i) non-electoral regimes, (ii) one- and no-party regimes, (iii) non-parliamentary constitutional monarchies, (iv) limited multi-party authoritarian

regimes, (v) exclusive democracies, (vi) male democracies, and (vii) electoral democracies. During the 2002-2014 period there were no countries coded as exclusive or male only democracies reflecting widespread near universal suffrage. In this paper we dichotomize the regime type variable into two categories: (1) democratic (lexical categories v-vii) and (ii) non-democratic (lexical categories i-iv). Example countries with their regime type classification in 2014 include Iran (non-democratic), Egypt (non-democratic), China (non-democratic), Russian Federation (non-democratic), Indonesia (democratic), Republic of Korea (democratic), and France (democratic). Our dichotomized variable corresponds, over the 2002-2014 period, to whether a country is an electoral democracy where "leaders are selected through contested elections held periodically before a broad electorate" [29, p. 1495].

As mentioned, we use the Freedom House press freedom measure.[5] In 2014, the measure included 197 countries. The press freedom score for each country is based upon answers from a panel of experts to 23 questions divided into categories of legal, political, and environmental environments. The result ultimately yields a country numeric score ranging between 0 (highest degree of press freedom) and 100 (lowest degree of press freedom). These numeric scores are then assigned categorical ones where a country's media is classified *Free* (numeric score less than 31), *Partly Free* (numeric score 31 to 60), or *Not Free* (numeric score 61 to 100). We use these categorical scores in our subsequent analyses. There are several reasons for this. Most importantly, our interest is in the impact of homophily and thus the numeric scores make little sense in that we would be forced to consider two countries with scores of 10 and 95 being, with respect to homophily, the same as if they were to have had scores of 10 and 11. That is, in neither case would they match since the numeric scores are not identical. Secondly, informal discussions with a rating panel participant suggests that numeric score changes at or near a category boundary receive considerable additional attention from rating panels. Thus we speculate that the categorical scores will be less noisy than the numeric ones.

Our focus is on understanding the underlying structure of direct Internet connections between countries and the extent to which both country-specific social and political variables and how countries relate with one another can account for that structure. To investigate this question we will examine several theoretically informed Exponential Random Graph (ERG) models accounting for the tie structure of the global Internet. Structure here is represented by a graph whose nodes are countries with at least one measurable Internet connection to at least one other country and whose edges are the direct Internet connections between all pairs of

---

2. Though our focus in this paper is on 2014, reference will also be made to our entire dataset which extends back to 2002.

3. *GNI* is gross domestic product plus incomes of foreign residents minus domestic earnings of non-residents. We use the World Bank's PPP, purchasing power parity, GNI to facilitate cross-national comparisons.

4. We consider domestic Internet users as a proxy for how accessible the Internet is to citizens. As a consequence, we normalize it by the total population. Thus, for example, we would not want large population countries where a small proportion of the population uses the Internet to look "better" than a much smaller country with a high proportion of its population using the Internet. *GNI*, on the other hand, we use as a proxy for the economic strength of a country and thus do not normalize it by population.

5. Freedom House's press freedom measure has been criticized both for being biased in favor of Western liberal values and for paying insufficient attention to new media such as blogs. Excellent summaries of these critiques are in [30] and [31].

countries in 2014.[6] More specifically, we assume the global Internet in a given year is comprised of a fixed set of n countries (nodes) together with the data links connecting them represented by the $n \times n$ adjacency matrix A where:

$$A_{i,j} = \begin{cases} 1, & \text{if } i, j \text{ share an Internet connection} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

Information can flow in either direction over an Internet connection and thus the graph is undirected and A is symmetric ($A_{i,j} = A_{j,i}$).[7] Our objective then is to estimate a stochastic model of A which also does a reasonable job accounting for other general properties of the observed Internet such as degree distribution, shared links, and geodesic distances.

ERG models, sometimes referred to as p-star models, assume that any given network is a realization of an underlying statistical process. In this sense, our 2014 network is the result of sampling from a distribution of networks each of which was generated by the same statistical model. Using ERG techniques is a several step process. First, we use existing data to generate a "best" approximation of that statistical process. We then use bootstrapping to assess how likely we are to generate simulated networks with macro network-level properties of the sort observed in our 2014 network assuming our observed dataset is not an extreme outlier in the actual underlying generating distribution.

What follows considers the 163-country subset for which we have data on all measured variables used in estimating the various models. This compares with the total of 201 countries with positive direct Internet bandwidth connections to at least one other country in 2014. Countries with no missing data comprise 94.24 percent of the total 2014 global bandwidth and 100 percent of all the edges (direct Internet connections) in the full dataset. The countries excluded from our analysis tend to be less connected (low degree) countries. Figure 1 displays the structure but with countries *not included* in our ERGM analyses shown in red along with their country names. Figure 2 illustrates the geographic distribution of press freedom status for the 163 in our 2014 analysis subset.

Previous work has shown that the evolution of the global Internet exhibits considerable path dependence associated with a country's economic and political situation [27], [32], [33]. These studies used country-level measures such as political regime-type, economic output, and domestic Internet penetration to account for variations in global bandwidth generally assuming that the attribute values for a given country are independent of those for other countries. We build on these past analyses by initially considering three country-level attribute variables–total global Internet bandwidth, gross national income (*GNI*), and domestic Internet

6. While we have 2015 data for most countries, global Internet bandwidth capacity data are often revised during the year following publication. Thus 2014 is the most recent year for which the measured values can be assumed reliable.

7. While we have bandwidth capacity for each data link, we are not using that information in this analysis and are treating the network as unweighted.

Figure 1. Global Internet Full Structure with Dropped Nodes in Red and Labelled: 2014

Figure 2. Freedom House Status (analysis country subset): 2014

users per 100 population. Given the heavy right tail of both total bandwidth and *GNI*, we use the natural log of these variables.

Our goal is to model whether any two countries share a direct Internet connection. To accomplish this, we go beyond node attribute values to explore the degree to which shared characteristics or homophily adds to our ability to predict direct Internet link between country pairs. So, for example, are two democratic countries more likely to share a direct global Internet connection, holding constant all other measured variables, than are democratic and non-democratic countries?

Direct Internet connections require capital investment and it seems reasonable to expect that countries near to one another may be more likely, again *ceteris paribus*, to share

a connection than would country pairs at a considerable geographic difference. We use continent as a proxy for this and expect continent homophily (country dyads in the same continent) to be predictive of direct connections.[8]

Regime type is another homophily variable. Are countries with identical regime types (democratic or non-democratic) more likely, *ceteris paribus*, to share a direct link? The Internet was based on packet switching research conducted in the U.S., France and Great Britain, all democracies, in the 1960s. This research led to the ARPANET in 1969 and by the 1980s inter-networking connected countries including Western European democracies, Australia, and, in Asia, first Japan and then South Korea. By 2002, the first year of our dataset, the global Internet connected 187 countries.

The early mover advantage in terms of connectedness as measured by the degree, number of direct global Internet ties with other countries for a given country, can be seen by comparing 2002 with 2014. Of the 174 countries for which we had regime type measures in 2002, 59.8 percent were classified as democratic. Importantly, the democratic countries were considerably more connected to the global Internet than were the non-democratic ones. The mean degree (number of direct connections) for democratic countries was 7.6 (median = 3) as compared to a mean degree for non-democratic countries of 3.2 (median = 2).[9]

As mentioned in the Background section, by 2014 the entire global Internet had become more decentralized with more countries having more direct data ties to other countries. Of the 182 countries for which we had regime type measures in 2014, 65.9 percent were classified as democratic, roughly the same as in 2002. The mean degree (number of direct connections) for democratic countries was 10.4 (median = 6) as compared to a mean degree for non-democratic countries of 8.1 (median = 5). While both democratic and non-democratic countries were, on average, more connected than in 2002, the differences had decreased. Thus we might expect regime type homophily to matter more in the earlier days of the global Internet than it would by 2014.

The final homophily variable we consider is press freedom. Are countries with the same press freedom categorization more likely, *ceteris paribus*, to share a direct connection? This might seem an odd expectation. Firstly, if the critique of the Freedom House measure that it simply reflects Western liberal values is correct, then we would anticipate regime type and press freedom to be highly confounded. However, it is interesting to look at Figure 3. This shows the network structure of our 2014 network. Here the nodes (countries) are colored as follows: blue border
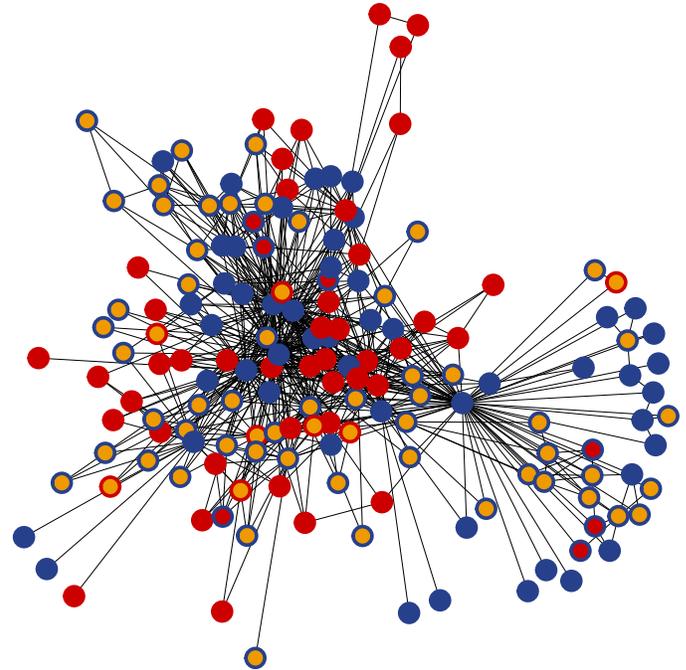

Figure 3. Global Internet Structure 2014 after Dropping Nodes (blue border indicates that the regime type is democratic, red border indicates the regime type is not democratic, blue interior color indicates press freedom status is free, red interior color indicates press freedom status is not free, and orange interior color indicates press freedom status is partly free)

indicates that the regime type is democratic, red border indicates the regime type is not democratic, blue interior color indicates press freedom status is free, red interior color indicates press freedom status is not free, and, finally, orange interior color indicates press freedom status is partly free. In trying to distinguish between regime type homophily and press freedom status homophily we are interested in how predictive border colors or interior colors are of an edge or link between any two countries.

To summarize, we examine whether plausible models of the tie structure of the global Internet can be estimated using $domestic\ Internet\ users\ per\ 100$, $log(global\ bandwidth)$, and $log(GNI)$ as attribute covariates and $regime\ type$, $UN\ continent$, and $press\ freedom\ status$ as homophily variables.

4. Results

Initially, two ERG models were estimated.[10] In all models, the dependent variable was whether or not a direct tie existed between country pairs. The main effects model for 2014 focused only on country attribute variables. The independent attribute variables for each country were (i) domestic Internet users per 100, (ii) total global bandwidth (logged), and (iii) GNI (logged). A summary of results is in

---

8. For this purpose we use the United Nations continent definitions. While UN region might seem a more fine grained measure, it fails to match countries sharing a land border while classified in different regions. An example would be the Czech Republic (Eastern European) and Austria (Western Europe).

9. The large differences between means and medians reflects the heavy right tail of the degree distribution. As an example, in 2002, the U.S. was the largest degree country with degree 124.

10. Models were estimated in *R* [34] using the *SNA* and *ergm* packages [10], [11].

Table 1. Since a tie (direct connection) either exists or does not exist between any two countries, the coefficients reflect the change in the log-odds likelihood of a direct connection for a unit change in the predictor variable. Standard errors are shown below the coefficients. In the main effects ERG model, the domestic Internet users per 100 was not statistically significant. However, both the GNI and bandwidth variables were statistically meaningful predictors.

All three attribute variables are treated as continuous as indicated by the `nodecov` preceding their names in Table 1. Using `nodecov` adds a statistic to the model equal to the sum of the attribute value for each dyad. For bandwidth the sum of the total global bandwidth (logged) for each pair of countries in the network would be added to the bandwidth statistic. This means that a large bandwidth-low bandwidth pair would look similar to a pair in which each country had moderate total bandwidth. It is also worth noting again that bandwidth here refers to each country's total global bandwidth and not to the shared bandwidth over the pair if there is a tie. Empirically this seems plausible in that we do tend to observe large bandwidth (or large GNI) countries connecting both to each other and to smaller ones. Basically, large bandwidth (or large GNI) countries tend also to be higher degree countries.

Table 1. 2014 MAIN EFFECTS ONLY

	2014 Network exponential family random graph
edges	−19.2000***
	(1.1500)
nodecov.UsersPer100	0.0021
	(0.0017)
nodecov.ln_BW	0.2540***
	(0.0307)
nodecov.ln_gni	0.1970***
	(0.0307)
AIC	4,362.0000
BIC	4,392.0000

*p < .05; **p < .01; ***p < .001

ERG model results are generally easier to interpret in terms of odds ratios which show any improvement a predictor variable provides within the estimated model. A coefficient is transformed to an odds ratio via an exponential transformation. So if the coefficient is $\hat{\theta}$, the odds ratio will be ($e^{\hat{\theta}}$). 95% confidence intervals are calculated from the coefficient's estimated standard error ($\hat{se}$).

$$95\%CI_{\hat{\theta}} = e^{\hat{\theta} \pm 1.96\hat{se}_{\hat{\theta}}} \qquad (2)$$

In what follows, we will summarize each model by providing odds ratios for the predictor variables using the 95% confidence interval. If a variable's odds ratio is above 1.0 throughout 95% confidence interval range, we say it significantly affects our ability to predict a direct Internet connection between two countries. In the case of the homophily variables, the odds ratio can be interpreted as indicating the change in odds of a direct tie assuming a match on the variable between two countries and all other variables held constant. Table 2 shows the information in Table 1 but now as odds ratios.

In this model, the bandwidth variable (odds ratio = 1.29 with lower 95% value = 1.21) and the *GNI* variable (odds ratio = 1.22 with lower 95% value = 1.15) were significant. The estimated odds ratio of 1.29 for bandwidth can be interpreted as saying that there is a 1.29 increase in the odds of a tie between two countries for every unit increase in (logged) global bandwidth.

Table 2. 2014 ODDS 95% CI MAIN EFFECTS MODEL

term	conf.low	estimate	conf.high
edges	0.00	0.00	0.00
nodecov.UsersPer100	1.00	1.00	1.01
nodecov.ln_BW	1.21	1.29	1.37
nodecov.ln_gni	1.15	1.22	1.29

The homophily model drops the Internet users per 100 country variable (since it was not a significant predictor in the main effects model) and adds as second order effects three homophily variables: (i) CONTINENT (to reflect that it may be easier for a country to connect to countries on the same continent), (ii) regime type (coded as *democratic* or *non-democratic*), and (iii) Freedom House status (coded as *free*, *partly free*, or *not free*). The three homophily variables are discrete and tested for exact value matching (*node.match* in model language) with odds results shown in Table 3. The country-level variables remained significant. Not surprisingly, of the homophily variables, CONTINENT ($p < .001$) had the highest estimated odds ratio indicating that knowing two countries are on the same continent increases the odds of their sharing a direct connection by a factor of a little over 4. Press freedom status homophily ($p < .01$) was also significant indicating that if two countries have the same press freedom status score, it increases the odds of a direct connection by 28% again assuming that all other variables are unchanged.[11] Importantly, the regime type variable was not significant.

It is clear that homophily with respect to CONTINENT is by far the most predictive followed by freedom house status homophily. Looked at in terms of probability, if two countries are both on the same continent and have the same press freedom status, the probability of their sharing a direct connection is 0.84 all other things being equal. If we only know that two countries have the same press freedom status then the probability of their sharing a direct connection is 0.56. And, finally, if all we know is that the countries are on the same continent, the probability of their sharing a connection is 0.81. These conditional probability estimates are obtained by doing an inverse logit transform on the estimated coefficients, $invlogit(\hat{\theta}) = \frac{1}{1+\exp(-\hat{\theta})}$.

Homophily means that connections form between countries who share values on the measured characteristic. A possibly confounding effect would be transitivity wherein countries who have connections in common will develop

11. As a point of comparison, the Las Vegas house advantage on roulette is estimated to be 5.26%.

direct connections with one another. To test for this we used the *(*ergm) term `gwesp` [35] to check for geometrically weighted edgewise shared partners by estimating a 2014 model using only `gwesp` and then the homophily model adding a `gwesp` term. While the `gwesp` term was statistically significant in both these models, the impact of both the covariates and the homophily terms remained predictive, within the 95% confidence interval, as in the base homophily model. We used .25 as the scaling parameter in the `gwesp` term. While we conclude that homophily matters, we did not identify an "optimal" scaling factor.[12]

Table 3. 2014 ODDS 95% CI HOMOPHILY MODEL

term	conf.low	estimate	conf.high
edges	0.00	0.00	0.00
nodecov.ln_BW	1.22	1.27	1.32
nodecov.ln_gni	1.20	1.26	1.32
nodematch.CONTINENT	3.56	4.23	5.03
nodematch.fh_status	1.06	1.28	1.55
nodematch.regime_type_binary	0.74	0.90	1.10

Comparing AIC scores, the homophily model (4071) is lower than that for the no homophily one (4362). Thus the homophily model is preferred and we consider its goodness of fit as shown in Figure 4. Here the bold line represents the observed data against data simulated from the estimated model. The results look fairly good for degree distribution though it underestimates the number of edge-wise shared partners. We also looked at goodness of fit with the `gwesp` term included. Results were similar, though AIC was lowered to 3794, and the number of edge-wise shared partners continued to be underestimated.

While the negative results for regime type may seem surprising, it is important to keep in mind that our results are with regard to the existence (or non-existence) of a direct tie between countries. When it comes to total global bandwidth, regime type clearly matters as shown in Table 4. The mean global bandwidth for democratic countries is almost four times greater than the mean for non-democracies.

To examine this further, we estimated the homophily model for each year in the 2002-2014 period of our dataset. The results, not shown, were that press freedom status was predictive in 2011, 2012, 2013, and 2014 but not before. The regime type variable was predictive of direct ties in 2002, 2003, and 2004. The country attribute variables remained predictive in every year for which we have data (2002-2014) as did the CONTINENT homophily variable. We have no plausible theoretical explanation for the recent significance of freedom of press variable. As far as we are aware, Freedom House made no methodological changes in its press freedom measure post 2002.[13] It is always a

12. The .25 is a typical value in the literature we have reviewed. Model estimation takes considerably more computational time with the `gwesp` term specified. We attempted to identify a "best" value for the `gwesp` term using tools provided within *ergm* however the estimate did not converge in over a week of run time on a 4 GHz iMac with 32GB of memory and increased `memory.size` allocation for *R*.

13. [30, p. 9] report that Freedom House made changes in its press freedom measure in 1989, 1994, 1997, 1999, and 2002.
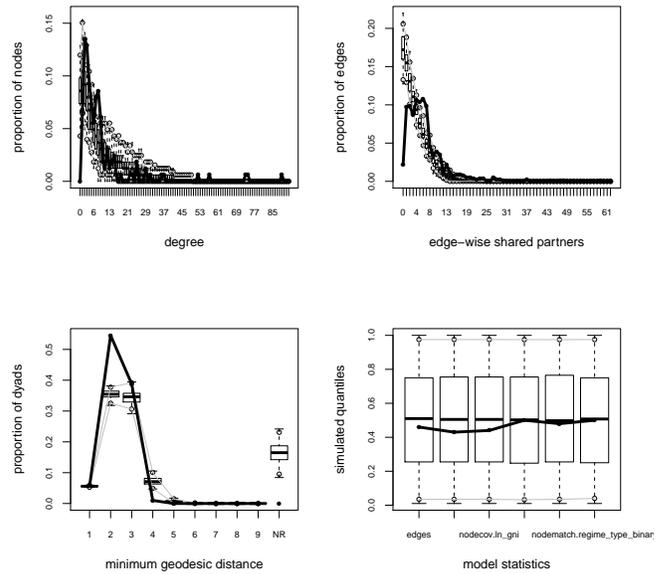

Figure 4. 2014 Homophily Model GOF

possibility that the data for any given year are anomalous in some senses. However, the fact that results seem patterned and consistent over the past four years argues against 2014 being an outlier. Finally, recall that ERGM is most effective when there is no missing data and our data set includes only those countries for which we have valid measures on all the model variables. A consequence is that 38, mostly smaller degree and bandwidth, countries were excluded. Countries with missing data in 2014 and thus not included can be seen in Figure 1.

Table 4. 2014 GLOBAL BANDWIDTH BY REGIME TYPE

regime-type	mean Tbps	median Tbps	total Tbps	n
Democratic	1.07	0.05	119.66	112
Not democratic	0.27	0.01	13.69	51

5. Discussion

Traditional news intermediaries including governments and mainstream media outlets are no longer exclusive gatekeepers as social media platforms including Twitter, Weibo, and Facebook permit people to share news and information directly with one another. That said, not everyone has easy access to the global Internet, a necessary condition for using social media. Country-level Internet infrastructure and policy play an important role in this, as demonstrated in North Korea's government policies denying Internet access to most of its citizens and the Chinese government generally restricting access to Google and most of its services. At the same time, a country's press freedom status provides an important snapshot of the country's legal, political, and economic conditions for citizens' access to information which

is increasingly available online. Even within countries of a similar regime type (e.g., democratic country), levels of press freedom can vary significantly.

Our results demonstrate that press freedom homophily is significantly predictive of direct Internet connections even in the context of geographic location and whether or not a country is democratic. That is, countries exhibiting homophily with regard to press freedom are more likely to share direct country-to-country Internet connections even after controlling for their economic performance, political regime type, and geographical location. Specifically, our study showed that press freedom status emerged as an important predictor of direct Internet ties between countries in 2011 and remained so until 2014, the final year of data analysis in this study. The regime type variable was a significant predictor in 2002, 2003, and 2004.

Our research offers several important scholarly and policy implications. First, this study provides important insights to understanding the underlying structure of direct Internet connections between countries. By analyzing regime type and press freedom along with other relevant variables such as geographic location and economic development, this study advances research on global Internet connectedness. In addition, our results may increase understanding of the relationship between press freedom and democracy. While these variables are often found to be highly correlated [22], in this paper we find (i) that the two variables play distinct roles in predicting the existence of direct global Internet connections and (ii) that the specific nature of these roles has changed over the 2002-2014 period.

Second, our analysis of press freedom should be informative for scholars who examine how press freedom might spread across the globe. This study suggests that more attention needs to be paid to associations between press freedom homophily and direct Internet ties between countries. Press freedom has rarely been examined in the context of global Internet connections, and thus our study fills a gap in the literature. Moreover, the one past study of press freedom contagion we identified [23] concluded that press freedom spread to nearby countries. In the age of the global Internet, "nearness" might be extended to countries sharing direct connections. If so, our positive results related to the impact of both edgewise shared partners and press freedom homophily appears consistent with a contagion effect. Consequently, our study helps scholars better understand the role of the Internet and press freedom in democratic changes.

Third, the ERG approach used in this study should be helpful for future research in related areas. Our study contributes to advancing computational methods in communication research by demonstrating how ERG approaches can be used to study democracy, press freedom, global Internet connectivity, and other related variables. ERGM has gained increased attention in the field of communication [4]–[6] but is still new to many scholars in the field.

Moreover, findings should be helpful for policymakers in the area of Internet connectivity and organizations that analyze and advocate for press freedom such as Freedom House and Reporters without Borders.

Future research should provide a more in-depth look at relationships between the variables each year. In addition, obtaining more complete datasets for countries is important though it probably will not be solved anytime soon. As mentioned before, missing data in our datasets prevented us from analyzing all the countries. While the countries excluded due to missing data tended to be low bandwidth and low degree, we cannot rule out that having complete measures on those countries would affect our results.

Finally, as with any modeling technique, ERGM can be extremely sensitive to what might appear to be fairly minor tweaks to the model being estimated [36]. This becomes especially noticeable when adding variables to the model to account for macro features such as shared edges, triangles, or degree. These and other model parameters can be adjusted to see if, say, AIC scores get lower. There is always the danger of overfitting by modeling noise in the data. In this paper we have kept the the models simple and made every effort to develop them in a manner informed by available literature.

References

[1] (2017, 02). [Online]. Available: https://www.bloomberg.com/politics/articles/2017-02-18/merkel-pushes-back-on-trump-s-media-attacks-calls-for-respect

[2] L. Kleinrock, "Information flow in large communication nets," *RLE Quarterly Progress Report*, vol. 1, 1961.

[3] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social networks*, vol. 1, no. 3, pp. 215–239, 1978.

[4] H. Song, "Uncovering the structural underpinnings of political discussion networks: Evidence from an exponential random graph model," *Journal of Communication*, vol. 65, no. 1, pp. 146–169, 2015.

[5] M. Shumate and E. T. Palazzolo, "Exponential random graph (p*) models as a method for social network analysis in communication research," *Communication Methods and Measures*, vol. 4, no. 4, pp. 341–371, 2010.

[6] B. F. Welles and N. Contractor, "Individual motivations and network effects a multilevel analysis of the structure of online social relationships," *The ANNALS of the American Academy of Political and Social Science*, vol. 659, no. 1, pp. 180–190, 2015.

[7] P. Pattison and S. Wasserman, "Logit models and logistic regressions for social networks: Ii. multivariate relations," *British Journal of Mathematical and Statistical Psychology*, vol. 52, no. 2, pp. 169–193, 1999.

[8] J. K. Harris, *An introduction to exponential random graph modeling*. Sage Publications, 2013, vol. 173.

[9] E. D. Kolaczyk and G. Csárdi, *Statistical analysis of network data with R*. Springer, 2014.

[10] C. T. Butts, *sna: Tools for Social Network Analysis*, 2016, r package version 2.4. [Online]. Available: https://CRAN.R-project.org/package=sna

[11] D. R. Hunter, M. S. Handcock, C. T. Butts, S. M. Goodreau, Morris, and Martina, "ergm: A package to fit, simulate and diagnose exponential-family models for networks," *Journal of Statistical Software*, vol. 24, no. 3, pp. 1–29, 2008.

[12] M. F. Guillén and S. L. Suárez, "Explaining the global digital divide: Economic, political and sociological drivers of cross-national internet use," *Social Forces*, vol. 84, no. 2, pp. 681–708, 2005.

[13] H. Seo and S. J. Thorson, "Networks of networks: Changing patterns in country bandwidth and centrality in global information infrastructure, 2002–2010," *Journal of Communication*, vol. 62, no. 2, pp. 345–358, 2012.

[14] C. Kedzie, "Communication and democracy: Coincident revolutions and the emergent dictators," 1997.

[15] H. V. Milner, "The digital divide the role of political institutions in technology diffusion," *Comparative Political Studies*, vol. 39, no. 2, pp. 176–199, 2006.

[16] P. Norris, *Digital divide: Civic engagement, information poverty, and the Internet worldwide*. Cambridge University Press, 2001.

[17] E. G. Rød and N. B. Weidmann, "Empowering activists or autocrats? the internet in authoritarian regimes," *Journal of Peace Research*, vol. 52, no. 3, pp. 338–351, 2015.

[18] Freedom House, "Freedom of the press, 2012," 2012.

[19] A. Alam and S. Z. Ali Shah, "The role of press freedom in economic development: A global perspective," *Journal of Media Economics*, vol. 26, no. 1, pp. 4–20, 2013.

[20] H. Tran, R. Mahmood, Y. Du, and A. Khrapavitski, "Linking global press freedom to development and culture: Implications from a comparative analysis," *International Journal of Communication*, vol. 5, p. 22, 2011.

[21] J. Whitten-Woodring, "Watchdog or lapdog? media freedom, regime type, and government respect for human rights," *International Studies Quarterly*, vol. 53, no. 3, pp. 595–625, 2009.

[22] K. D. Karlekar and L. B. Becker, "By the numbers: Tracing the statistical correlation between press freedom and democracy," 2014.

[23] R. S. Sobel, N. Dutta, and S. Roy, "Beyond borders: Is media freedom contagious?" *Kyklos*, vol. 63, no. 1, pp. 133–143, 2010.

[24] World Bank, "World development indicators," http://data.worldbank.org/data-catalog/world-development-indicators, July 2016, accessed: 2016-07-08. [Online]. Available: http://data.worldbank.org/data-catalog/world-development-indicators

[25] Institut for Statskundskab, "Lexical index of electoral democracy," http://ps.au.dk/forskning/forskningsprojekter/dedere/datasets/, July 2016, accessed: 2016-07-06. [Online]. Available: http://ps.au.dk/forskning/forskningsprojekter/dedere/datasets/

[26] TeleGeography, "Global Internet geography," PriMetrica, Inc., 2016.

[27] G. A. Barnett and H. W. Park, "The structure of international internet hyperlinks and bilateral bandwidth," *Annals of Telecommunications*, vol. 60, no. 9, pp. 1110–1127, 2005.

[28] Freedom House, "Freedom in the world 2015," *interactive map, online: https://freedomhouse. org/report/freedom-world/freedom-world-2015*, 2015.

[29] S.-E. Skaaning, J. Gerring, and H. Bartusevičius, "A lexical index of electoral democracy," *Comparative Political Studies*, p. 0010414015581050, 2015.

[30] J. Burgess, *Evaluating the evaluators: media freedom indexes and what they measure*. Center for International Media Assistance, National Endowment for Democracy, 2010.

[31] L. B. Becker, T. Vlad, and N. Nusser, "An evaluation of press freedom indicators," *International Communication Gazette*, vol. 69, no. 1, pp. 5–28, 2007.

[32] G. A. Barnett, "A longitudinal analysis of the international telecommunication network, 1978-1996," *American Behavioral Scientist*, vol. 44, no. 10, pp. 1638–1655, 2001.

[33] H. Seo and S. Thorson, "A mixture model of global Internet capacity distributions," *Journal of the Association for Information Science and Technology*, 2015.

[34] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2016. [Online]. Available: https://www.R-project.org/

[35] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, "An introduction to exponential random graph (p*) models for social networks," *Social networks*, vol. 29, no. 2, pp. 173–191, 2007.

[36] C. R. Shalizi and A. Rinaldo, "Consistency under sampling of exponential random graph models," *Annals of statistics*, vol. 41, no. 2, p. 508, 2013.