Explaining the Relationship between Internet and Democracy in Partly Free Countries Using Machine Learning Models

MUSTAFA SAGIR, Department of Political Science, University of Cincinnati, United States
SAID VARLIOGLU, School of Information Technology, University of Cincinnati, United States

ABSTRACT
Previous studies have offered a variety of explanations on the relationship between democracy and the internet. However, most of these studies concentrate on regions, specific states or authoritarian regimes. No study has investigated the influence of the internet in partly free countries defined by the Freedom House. Moreover, very little is known about the effects of online censorship on the development, stagnation, or decline of democracy. Drawing upon the International Telecommunication Union, Freedom House, and World Bank databases and using machine learning methods, this study sheds new light on the effects of the internet on democratization in partly free countries. The findings suggest that internet penetration and online censorship both have a negative impact on democracy scores and the internet’s effect on democracy scores is conditioned by online censorship. Moreover, results from random forest suggest that online censorship is the most important variable followed by governance index and education on democracy scores. The comparison of the various machine learning models reveals that the best predicting model is the 175-tree random forest model which has 92% accuracy. Also, this study might help "IT professionals" to see their important role not only in the technical fields but also in society in terms of democratization and how close IT is to social sciences.

CCS Concepts: • Social and professional topics → Geographic characteristics.

1 INTRODUCTION
In a speech at the Paul H. Nitze School of Advanced International Studies of the Johns Hopkins University, March 9, 2000, on China W.T.O. agreement, President Clinton said: "In the new century, liberty will spread by cell phone and cable modem. Now there is no question China has been trying to crack down on the Internet. Good luck! That’s sort of like trying to nail Jell-O to the wall" [7]. This perspective shows the clear optimistic view of the internet’s prospective influence on democratization in the early 2000s. However, leaders are more cautious about this optimistic prospect with the different developments in the following years. First of all, internet penetration has kept on increasing in every country in the world. Especially, with the rise of smartphones, more people have access to the internet and people spend more time on the internet. International Telecommunication Unit reports that in 2019, 53.6 percent of the global population, are using the Internet. This percentage was only 16.8 percent in 2005 [19]. Also, social media usage has...
significantly increased. For instance, 85 percent of the adult population uses the Internet regularly and 80 percent of those people also use Facebook in the US [10]. The use of the internet and social media in the protests around the world strengthened the arguments that the internet can help democratization by providing an environment for sharing information and organizing.

On the other hand, Freedom House reported in 2019 that democracy scores have kept declining since 2005 [15]. This 13 year of decrease is not limited to some regions or only some categories. Countries from free, partly free, and not free countries and countries from all regions have lower scores in 2018 than in 2005 [14]. Thus, after the third wave of democratization which started with the Carnation Revolution in Portugal in 1974 and accelerated after the end of the Cold War, a backward course has been going on.

After the optimistic statements of the western leaders, we have seen that authoritarian regimes have used the internet to track the political activist, prosecute the dissidents, and spread their propaganda. Huntington argues “democracies are created not by causes but by causers” [18]. Drawing upon the same analogy, autocracies are also created by causers and according to one point of view, the internet gave more opportunities for authoritarian leaders to protect and establish their regimes. For example, digital rights group Access Now’s 2018 report reveals that in 2018 the internet is shut down 196 times at a national or regional level more than 90 times in 2017 [28].

Moreover, the internet as a double-edged sword can help share news and information but it can also be used to spread inaccurate information, manipulate voters, increase the polarization and hatred toward other people.

A variety of studies investigated the relationship between the internet and democracy. Some studies found a positive association, but some studies which especially focus on case studies in authoritarian regimes found a negative effect of the internet on democratization. Also, some studies explored the effects of diffusion of false news through the internet and especially via social media. However, most of these studies concentrate on regions, specific states or authoritarian regimes. Freedom House puts countries in three categories; free, partly free, and not free. According to Freedom House 2019 Report [14], 63 out of 205 countries are in the partly free category in 2019, but no study has explored the impact of the internet in these countries. Moreover, Freedom House prepares the “Censorship on Net” reports since 2013 [16], but no study has investigated the relationship between internet access, online censorship, and democratization.

The purpose of this quantitative study is to explore the influence of internet penetration and online censorship on democratization controlling economic, educational, life quality, and state capacity variables. Drawing upon International Telecommunication Union (ITU) internet penetration data [19], Freedom House Democracy and internet censorship scores [14], World Bank World Development Index [2], and World Bank Governance Index data for 2013-2017 [3], this study sheds light on the effects of the internet and online censorship on democratization in partly free countries. The current study applies multivariate regression analysis and machine learning models to better understand the importance of variables affecting the democracy scores. Thus, there are several key areas where this study makes an original contribution to this growing area of research.

2 THEORETICAL PERSPECTIVES AND LITERATURE

There is a large volume of published studies describing the role of several factors in promoting, accelerating or preserving democratization. Moreover, in parallel with the people’s access to the internet around the world, a large and growing body of literature has investigated the relationship between the internet and democracy. These studies can be broadly categorized under three distinct groupings. The first group of studies argues that with free access to information, knowledge sharing without any constraint, and the spread of political knowledge, the internet changes people’s political
attitudes and they found a positive effect of the internet on democratization [1, 6, 21, 22, 24, 26, 27, 29].

The second group of studies that especially focus on authoritarian regimes argues that the internet gives a distinct means to track the political activists, prosecute the dissidents, and spread the regime propaganda, which in turn limits the democratization [11, 12, 20, 25, 30].

The third group of studies explored the effects of diffusion of false news through the internet and especially via social media. These studies argue that the internet and especially social media is not beneficial but detrimental to democracy [4, 5, 8, 9, 17, 23].

Although an extensive body of research has been carried out and studies over the past two decades have provided valuable information on the relationship between internet/social media use and several aspects of democracy, no single study exists which specifically focuses on partly free countries. Moreover, little is known about the effects of online censorship on the development, stagnation, or decline of democracy.

3 THE CURRENT STUDY

This quantitative study tries to fill these gaps by analyzing the relationship between internet penetration and democracy in partly free countries. This study also investigates the interaction between internet penetration and online censorship to see whether the democracy scores of partly free countries are conditioned by the level of online censorship that exists in those states.

3.1 Research Question and Hypotheses

The primary aim of the study is to examine the effects of internet penetration and online censorship on democratization in partly free countries controlling the economic, educational, and life quality variables. Thus, the central research question is: “To what extent does internet promote, consolidate, and protect democracy in partly free countries?”

As outlined in the literature review section, there are contradicting arguments on the effects of the internet on democratization. Consequently, there are competing hypotheses and expectations about how the increase in the usage of the internet may affect democratization.

Drawing on the argument that the internet accelerates the democratization, Hypothesis 1: An increase in the accessibility of the internet increases the democracy scores in partly free countries.

The second argument is that governments can use censorship tactics to prevent information sharing, track dissidents, and block social unrest by using the internet, which in turn decline democratization. Hence, Hypothesis 2: An increase in internet censorship prevents democratization and decrease democracy scores.

In the third model, we investigate the moderation effect of online censorship. Accordingly, Hypothesis 3: Internet penetration’s effect on democratization is conditioned by online censorship.

3.2 Dependent Variable

The dependent variable in the current study is “Freedom House Total Aggregate Scores” for the period 2013-2017. Freedom House assesses Political Rights in 10 indicators and Civil Liberties in 15 indicators. Each indicator has calculated from 0-4. Thus, each country or territory has a Political Right score of 0-40, Civil Liberties score of 0-60, and Total Aggregate scores of 0-100. This study uses the Freedom House Total Aggregate scores as the dependent variable.

3.3 Independent Variables

The primary predictors of interest in the study are internet penetration and online censorship. Internet penetration is measured by the share of the population using the internet and obtained from the ITU website [19]. Online censorship is measured by Freedom House [14] with a methodology.
that includes 21 areas in three categories. These categories are Obstacles to Access, Limits on Content, and Violations of User Rights. Freedom on Net scores is the measure of how governments and non-state actors around the world limit the fundamental rights online. Each country has a score of 0-100, 0 being the most restrictive and 100 being the most free.

3.4 Control Variables
There is a vast literature attempting to explain the variables influencing the emergence and the survivability of democracies. Drawing on these previous studies, this study employs variables from the World Bank World Development Indicators Database and Worldwide Governance Indicators Database. The variables related to economy, education, life quality, urbanization, and demographics were obtained from World Bank World Development Indicators Database These are; GDP per Capita, Annual percentage of GDP per Capita Growth, Unemployment percentage of total population, Foreign Direct Investment percentage of GDP, Inflation, Government Expenditure on Education percentage of GDP, Military Expenditure percentage of GDP, Life Expectancy, Number of Infant Deaths, Urban Population, and Annual percentage of Urban Population Growth. Since there are no consistent data on Educational Attainment, Income Distribution, and Share of the Middle Class, these variables are omitted from the model.

We also included the variables related to state institutional capacity from five of the six World Bank Worldwide Governance Index elements of governance; Political Stability and Absence of Violence, Government Effectiveness, Regulatory Quality, Rule of Law, and Control of Corruption. The scores for these variables are range from -2.5 to 2.5, with greater values equivalent to better governance.

3.5 Analytic Approach
To be able to assess the effects of internet penetration and online censorship on democracy scores, first, we looked at the descriptive analysis to see how the variables are distributed across countries. Second, we investigated the bivariate analysis to see the correlation between variables. After finding that some variables are highly correlated with each other, we created a state capacity index which is composed of five Worldwide Governance Index’s five elements of governance. As a next step, we moved to multivariate analysis to be able to analyze the relationships among all variables to get a true picture. We scaled all the independent variables and ran the OLS Regression. Then, we created an interaction term between internet penetration and online censorship to see if the internet’s effect on democracy scores is conditioned by the level of online censorship in the countries in the sample and ran the OLS Regression again. After the regression analysis, we conducted decision tree, SVR, and random forest machine learning models analysis to find which variables are more important in the causality for democracy scores. We compared the test set results with actual results. Random Forest model has higher accuracy in predicting the democracy scores.

In addition, we conducted diagnostic tests to ensure that assumptions are met. We checked the multicollinearity and outliers in the data by careful selection of predictors and combining related measures into scales so that the collinearity issue was reduced and did not seem to be a problem.

4 FINDINGS
4.1 Descriptive Findings
For the 5-year period during 2013-2017, Freedom House Aggregate scores range from 25 to 79 in the countries which were in defined as partly free countries. The internet penetration levels and online censorship scores are also greatly varied and are ranging from 5.05-84.45 for the internet penetration and 24-71 for the censorship scores. Venezuela and Turkey are the countries that
showed high levels of decline in democracy scores and an increase in online censorship scores while Tunisia is the country that showed improvements in many areas.

4.2 Bivariate Relationship

The bivariate correlations shown in Table 1 demonstrate the baseline associations between internet penetration, online censorship, various development, and governance index variables and the dependent variable Freedom House Aggregate Scores. Internet penetration and FH scores have a positive correlation, but it is not statistically significant. However, there is a statistically significant negative correlation between online censorship and FH scores. Moreover, there is a statistically significant positive correlation between democracy scores and GDP per capita growth, Foreign Direct Investment, and Governance Index.

4.3 Multivariate Relationship

As a next step, we moved to multivariate analysis to be able to get a true picture. Table 2 shows the results of regression models. In the first model, just internet penetration was applied as an independent variable. In the second model, censorship scores are put in the model and internet penetration was removed from the model. In the third model, both variables are included and in the fourth model, the interaction term between internet penetration and online censorship was included to see whether internet penetration’s effect on Freedom House scores is conditioned by online censorship.

According to regression analysis, both internet penetration and online censorship have a negative association with democracy scores. Moreover, the interaction term in the fourth model shows that the internet’s effect on democracy scores is conditioned by online censorship.

4.4 Decision Tree and Random Forest

After the regression analysis, we conducted decision tree, SVR, and random forest machine learning models analysis to find which model has higher accuracy in predicting the democracy scores and which variables are more important in the causality for democracy scores. Then, we compared the test set results with actual results. Thus, we started with the decision tree. Figure 1 shows the decision tree. As seen in the picture, the machine first goes to censorship, then GDP growth and population, then life expectancy, foreign direct investment, and urban population growth according to the results of the previous inquiry.
Table 2. Multivariate Analysis

| VARIABLES                        | FH Scores (1) | FH Scores (2) | FH Scores (3) | FH Scores (4) |
|----------------------------------|---------------|---------------|---------------|---------------|
| Year                             | 1.35          | 0.42          | 2.93**        | 2.96**        |
|                                  | (0.97)        | (0.79)        | (0.86)        | (0.84)        |
| Internet Penetration             | -5.45**       | -8.61***      | -7.94***      | -8.60***      |
|                                  | (2.32)        | (2.28)        | (2.25)        | (2.28)        |
| Online Censorship                | -9.17***      | -6.13***      | -6.82***      | -7.45***      |
|                                  | (1.09)        | (1.06)        | (1.05)        | (1.07)        |
| GDP Per Capita                   | -5.84***      | -2.47         | -2.88         | -3.01         |
|                                  | (2.04)        | (1.97)        | (1.87)        | (1.93)        |
| GDP Per Capita Growth            | -1.39         | -2.24*        | -1.73         | -1.58         |
|                                  | (1.27)        | (1.18)        | (1.12)        | (1.10)        |
| Inflation                        | -2.50**       | -1.18         | -1.56         | -0.46         |
|                                  | (1.44)        | (1.35)        | (1.28)        | (1.23)        |
| Military Expenditure (% GDP)     | -5.62**       | -2.94**       | -3.96***      | -4.77***      |
|                                  | (1.30)        | (1.40)        | (1.35)        | (1.36)        |
| Government Expenditure on Education | 0.74         | 0.57          | 0.06          | 0.39          |
|                                  | (1.12)        | (1.04)        | (0.99)        | (0.99)        |
| Life Expectancy                  | 6.24***       | 4.33***       | 6.71***       | 6.63***       |
|                                  | (2.09)        | (1.83)        | (1.83)        | (1.84)        |
| Number of Infant Deaths          | 2.69          | -0.16         | 1.28          | 0.45          |
|                                  | (1.75)        | (1.59)        | (1.56)        | (1.56)        |
| Population Ages15-64              | 0.38          | 4.22***       | 1.53          | 1.28          |
|                                  | (1.71)        | (1.42)        | (1.52)        | (1.49)        |
| Population Growth                | -4.07         | -2.89         | -0.89         | -1.60         |
|                                  | (2.69)        | (2.51)        | (2.43)        | (2.39)        |
| Urban Population                 | 6.89***       | 0.30          | 6.51***       | 6.59***       |
|                                  | (2.50)        | (1.55)        | (2.21)        | (2.16)        |
| Urban Population Growth          | 4.32          | 3.36          | 0.28          | 3.17          |
|                                  | (2.75)        | (2.52)        | (2.52)        | (2.72)        |
| Female Population                | 4.39***       | 2.37          | 1.93          | 2.49*         |
|                                  | (1.57)        | (1.53)        | (1.45)        | (1.43)        |
| Governance Index                 | 5.59***       | 3.68***       | 3.68***       | 3.17          |
|                                  | (1.94)        | (1.84)        | (1.74)        | (1.71)        |
| Constant                         | 50.73***      | 50.73***      | 50.73***      | 50.32***      |
|                                  | (0.83)        | (0.77)        | (0.73)        | (0.73)        |

Standard errors in parentheses, ***, p<0.01, ** p<.05, * p<0.1

Fig. 1. Decision Tree

Then we moved to random forest. After trials of a different number of trees in Random Forest, the 175-tree model gave greater accuracy in predicting the Freedom House aggregate scores. We compared the prediction results in each of the models; SVR, decision tree and various combinations of random forests. Random Forest with 175 tree gave the best prediction accuracy. We compared the predictions of each model. According to our comparison, the prediction accuracy of the regression and the decision tree models are close, 84.6 percent and 84.9, respectively. 500-tree random forest’s prediction accuracy is 87.2 percent and 175-tree random forest’s prediction accuracy is 91.9 percent.
Furthermore, Figure 2 shows the dot chart of the importance of variables based on a 175-tree random forest. As seen in the figure, online censorship is the most important variable in the model whereas internet penetration is not a very significant variable.

![Fig. 2. Dot chart of variable importance scores based on Random Forest with 175 trees.](image)

Overall, these results indicate that there is a strong negative impact of online censorship on democracy scores whereas internet penetration’s influence is more mixed. Also, consistent with the earlier studies and theories, state institutional capacity and urban population have a positive association while military expenditure percentage of GDP has a negative association with democratization. Moreover, 175-tree random forest models’ predictions are more accurate than those of a decision tree or regression models. Together these results provide important insights into internet-democracy relationship literature.

### 4.5 Discussion and Conclusion

This study certainly adds to our understanding of the internet democracy relationship and provides a different perspective. The current study aimed to determine the effects of internet penetration and online censorship on democracy scores in partly free countries. The most obvious finding to emerge from this study is that online censorship has a strong negative impact on democratization in the studied countries (Hypothesis 1). However, we are cautious to have a casual argument here because we could not determine which direction the causality runs. It is also possible that countries use online censorship as a tool when they become more anti-democratic. For instance, Hellmeier claims that monarchies, regimes with greater levels of protests, and less opposition in the political field are more likely to censor the Internet[13].

In addition to the online censorship’s impact, internet penetration, overall, has a negative effect on democratization (Hypothesis 2), however, the partial dependence plots in Figure 3 after the random forest analysis reveal that the internet’s effect is more complicated and similar to a reverse U shape. Another significant finding to emerge from this study is that the internet’s effect on democratization is conditioned by online censorship (Hypothesis 3). Findings also suggest that state education, and urban population, and a state’s institutional capacity as measured by governance index are among the variables which have a positive association with democracy scores. On the other hand, an increase in the military expenditure percentage of GDP decreases the democratization.
Moreover, with the machine learning models, by using the independent and control variables, we can predict the democracy scores of partly free countries with a prediction accuracy between 84 and 92 percent. On average, random forest with 175-tree gave the best prediction accuracy with 92 percent accuracy.

Fig. 3. The Partial Dependence Plots

These findings have significant implications for the understanding of how online censorship influence the democratization in partly free countries. The present study is one of the first attempts to explore this relationship. The findings of the study will lay the groundwork for future research and help to improve explanations of the internet democracy relationship studies. Further research, especially on the online censorship-democracy relationship, would help us to establish a higher degree of accuracy on this issue.

REFERENCES

[1] Catie Snow Bailard. 2012. A field experiment on the internet’s effect in an African election: savvier citizens, disaffected voters, or both? Journal of Communication 62, 2 (2012), 330–344.

[2] World Bank. 2019. World Development Index Project. World Bank Documents (2019). http://datatopics.worldbank.org/world-development-indicators/

[3] World Bank. 2019. Worldwide Governance Indicators (WGI) Project. World Bank Documents (2019). https://info.worldbank.org/governance/wgi/

[4] Jamie Bartlett. 2018. The people vs tech: How the internet is killing democracy (and how we save it). Random House.

[5] Samantha Bradshaw and Philip N Howard. 2018. Challenging truth and trust: A global inventory of organized social media manipulation. The Computational Propaganda Project (2018).

[6] Anita Breuer and Jacob Groshek. 2014. Online media and offline empowerment in post-rebellion Tunisia: An analysis of Internet use during democratic transition. Journal of Information Technology & Politics 11, 1 (2014), 25–44.

[7] Bill Clinton. 2000. Speech on China Trade Bill. The New York Times (2000).

[8] Anamitra Deb, Stacy Donohue, and Tom Glaisyer. 2017. Is Social Media a Threat to Democracy? (2017).

[9] Ronald J Deibert. 2019. The road to digital unfreedom: Three painful truths about social media. Journal of Democracy 30, 1 (2019), 25–39.

[10] Shannon Greenwood, Andrew Perrin, and Maeve Duggan. 2016. Social media update 2016. Pew Research Center 11, 2 (2016).
[11] James Griffiths. 2019. *The great firewall of China: How to build and control an alternative version of the internet*. Zed Books Ltd.

[12] Jacob Groshek and Kate Mays. 2017. A time-series, multinational analysis of democratic forecasts and emerging media diffusion, 1994–2014. *International Journal of Communication* 11 (2017), 23.

[13] Sebastian Hellmeier. 2016. The dictator’s digital toolkit: Explaining variation in Internet filtering in authoritarian regimes. *Politics & Policy* 44, 6 (2016), 1158–1191.

[14] Freedom House. 2018. Freedom in the World 2018. (2018). https://freedomhouse.org/sites/default/files/FH_FITW_Report_2018_Final_SinglePage.pdf

[15] Freedom House. 2019. Freedom in the World 2019. (2019). https://freedomhouse.org/sites/default/files/Feb2019_FH_FITW_2019_Report_ForWeb-compressed.pdf

[16] Freedom House. 2019. FREEDOM ON THE NET 2019. (2019). https://freedomhouse.org/sites/default/files/2019-11/11042019_Report_FH_FOTN_2019_final_Public_Download.pdf

[17] Philip N Howard and SC Woolley. 2016. Political communication, computational propaganda, and autonomous agents-Introduction. *International Journal of Communication* 10, 2016 (2016).

[18] Samuel Huntington. 1968. Political order in changing societies New Haven: Yale U. (1968).

[19] ITU. 2019. Measuring digital development: Facts and figures 2019. *ITU’s Measuring digital development* (2019). https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx

[20] Shanthi Kalathil and Taylor C Boas. 2001. The Internet and state control in authoritarian regimes: China, Cuba and the counterrevolution. *First Monday* 6, 8 (2001).

[21] Ya-Wen Lei. 2011. The political consequences of the rise of the Internet: Political beliefs and practices of Chinese netizens. *Political Communication* 28, 3 (2011), 291–322.

[22] Tansa George Massoud, John A Doces, and Christopher Magee. 2019. Protests and the Arab Spring: An Empirical Investigation. *Politics* 51, 3 (2019), 429–465.

[23] Robert W McChesney. 2016. *Rich media, poor democracy: Communication politics in dubious times*. New Press, The.

[24] Ali Pirannejad. 2017. Can the internet promote democracy? A cross-country study based on dynamic panel data models. *Information Technology for Development* 23, 2 (2017), 281–295.

[25] Espen Geelmuyden Rød and Nils B Weidmann. 2015. Empowering activists or autocrats? The Internet in authoritarian regimes. *Journal of Peace Research* 52, 3 (2015), 338–351.

[26] Jaime E Settle, Robert M Bond, Lorenzo Coviello, Christopher J Fariss, James H Fowler, and Jason J Jones. 2016. From posting to voting: The effects of political competition on online political engagement. *Political Science Research and Methods* 4, 2 (2016), 361–378.

[27] Elizabeth Stoycheff, Erik C Nisbet, and Dmitry Epstein. 2016. Differential effects of capital-enhancing and recreational Internet use on citizens’ demand for democracy. *Communication Research* (2016), 0093650216644645.

[28] Berhan Taye, Sage Cheng, and →. 2019. The State of Internet shutdowns in 2018. https://www.accessnow.org/the-state-of-internet-shutdowns-in-2018/

[29] Leizhen Zang, Feng Xiong, and Yanyan Gao. 2019. Reversing the U: New Evidence on the Internet and Democracy Relationship. *Social Science Computer Review* 37, 3 (2019), 295–314.

[30] Zhi-Jin Zhong, Tongchen Wang, and Minting Huang. 2017. Does the Great Fire Wall cause self-censorship? The effects of perceived internet regulation and the justification of regulation. *Internet Research* (2017).