

Online ethnic segregation in a post-conflict setting

Annerose Nisser and Nils B Weidmann 

University of Konstanz, Germany

Abstract

Existing research has shown that online networks are often segregated along identity lines, such as political ideology or religious views. Although online segregation should be specifically detrimental when appearing between ethnic groups in a post-conflict setting, to date we have no systematic evidence on the level of online ethnic segregation. To close this gap, the present study examines online ethnic segregation in a large ethnically mixed blogger network in a post-conflict society, Bosnia and Herzegovina. Since politics has been found to enhance ethnic divides in the offline world, we additionally examine whether segregation is higher for bloggers engaging with political topics. Using large-scale web scraping, automated text analysis and Monte Carlo simulation, we find evidence for pronounced ethnic divisions. Furthermore, we find that political bloggers tend to have more ethnically segregated networks. The findings show that a broad public exchange transcending ethnic categories remains limited in the online context we study, and that those who dominate the online political debate tend to be those who in their social interactions put even more weight on ethnic categories than the average.

Keywords

Blogging, ethnicity, homophily, online fragmentation, online segregation

Introduction

Existing research has shown that online networks are often segregated: they cluster along political ideology (Adamic and Glance, 2005; Barberá et al., 2015; Pariser, 2011; Tremayne et al., 2006), national lines (Etling et al., 2010) or religious views (Kelly and Etling, 2008). While the concept of segregation has been called differently (online ghettos, online enclaves, echo chamber, online homophily, audience fragmentation, selective

Corresponding author:

Nils B Weidmann, Department of Politics and Public Administration, University of Konstanz,
Konstanz 78457, Germany.

Email: nils.weidmann@uni-konstanz.de

exposure or filter bubble), the phenomenon is generally assumed to be detrimental to an inclusive public debate, and to contribute to the polarization of the larger public (Barberá et al., 2015; Gitlin, 1998; Prior, 2007; Sunstein, 2001). El-Bermawy (2016) has even argued that ‘the filter bubble is destroying democracy’. While this claim may be exaggerated, we have good reason to be concerned about online segregation: a lack of contact between individuals of different groups (Allport, 1954) and a lack of contact with conflicting opinions (Mutz, 2002) have been associated with lower tolerance towards other groups and other opinions.

The existence of online segregation should be of even greater concern in countries with deep-rooted ethnic cleavages, such as countries that have experienced violent conflict in the past. In such contexts, political and societal polarization resulting from online echo chambers could have very tangible consequences. Yet, online fragmentation has, to date, been mainly studied in relatively ‘tame’ political contexts, such as cleavages between Democratic and Republican bloggers in the United States (e.g. Adamic and Glance, 2005; Barberá et al., 2015; Colleoni et al., 2014; Tremayne et al., 2006). In addition, existing research has not yet examined whether politics exacerbates online fragmentation, although politics frequently constitutes a contested and emotionally charged issue. To close this gap, the present study investigates online fragmentation between ethnic groups in a multi-ethnic, post-conflict society, and examines whether engagement with and active interest in politics increases online ethnic fragmentation.¹ By ethnic groups, we understand groups defined by a (belief in a) shared history, common descent, language and/or phenotypical features (Chandra, 2006; Vogt, 2014). By definition, ethnic identities are less malleable and more easily visible than most other identities (Chandra, 2006). We focus on the case of a large blogger network in Bosnia and Herzegovina. Bosnia and Herzegovina experienced violent conflict between ethnic groups in the 1990s, and the ethnic cleavages central to this conflict are still dominant in today’s political life. We expect that offline segregation and ethnic polarization transcend into the online sphere, and that bloggers become ethnically segregated despite the facts that individuals’ ethnicity is more difficult to identify and geographical separation is overcome online. Furthermore, we expect that engagement with politics increases the level of ethnic segregation in bloggers’ personal networks. For the offline world, it has been argued that ethnicity gains cultural and social importance when political actors ‘play the ethnic card’ as a strategy to gain support. We argue that politics in a post-conflict society is a sensitive, contentious and emotionally charged issue. In line with the argument that threat and contentiousness increase individuals’ search for consonant information (Lavine et al., 2005), we expect that politically engaged bloggers are more ethnically segregated than other bloggers. Our findings support this argument: bloggers in Bosnia and Herzegovina are highly ethnically segregated, and the existing level of segregation is not explained by any sociodemographic or network factors except ethnicity. Moreover, we find indeed that bloggers writing on political topics establish more ethnically segregated networks than other bloggers.

Our findings suggest that the fragmentation of information exchange and communication along ethnic lines cannot purely be explained by ethnically segregated institutions in the offline world but continue to exist in an environment where social control and physical boundaries are largely absent, and ethnic categories can exert influence only through

subliminal cues. Furthermore, our findings suggest that engagement with politics indeed impacts the ethnic composition of bloggers' personal networks. This implies that politics as an often contested, emotionally and ethnically loaded issue influences the information environment an individual chooses. In other words, those who actively engage in the online political debate tend to be ethnically more segregated than those who engage in other online debates, such as discussions around sports or fashion.

The study proceeds as follows: we first present the existing literature on ethnic segregation and the impact of politics on inter-ethnic relations, and introduce our hypotheses (see 'Theory and hypotheses' section). We then describe the empirical case examined in our study, as well as language peculiarities of the Bosnian context important for carrying out our study. We also offer an overview of the coding of key variables and the research design (see 'Case selection and methods' section). Finally, the 'Results' section presents our findings. In the last section of this article, we discuss the implications of our findings, as well as next steps to take in future research.

Theory and hypotheses

Online networks have been found to be fragmented around political ideology, especially the Democrat–Republican divide in the United States (Adamic and Glance, 2005; Barberá et al., 2015; Colleoni et al., 2014; Tremayne et al., 2006), national lines (Etling et al., 2010) and religious views (Kelly and Etling, 2008). By online *fragmentation* or online *segregation*, we refer to the fact that people preferentially connect and exchange information with others who are like-minded or similar. Although this phenomenon even occurs in offline contexts, there are two reasons why it is of specific relevance online. First, the Internet is a 'high-choice environment' (Garrett, 2009) where individuals are free to choose their sources of information and their interaction partners from a large set of available options. While people's interaction partners are often set by the environment in offline contexts, people have an easier time to self-select into segregated, homogeneous environments online. Second, algorithms on search portals and social networking sites are optimized to show content that individuals tend to like, and this is most often content similar to what individuals have seen before (Ebo, 1998; Pariser, 2011). To summarize, people's tendencies of choosing similar or like-minded interaction partners can be enhanced online by (1) the large set of available options and (2) by the design of online services. For these reasons, Morozov (2012) speaks of the 'Splinternet', and Ebo (1998) speaks of 'cyberghettos'.

But why should we care about the phenomenon of online segregation? Lately, the phenomenon has been discussed in the context of elections, especially the 2016 election of President Trump in the United States (e.g. Wong et al., 2016). It has been argued that online segregation can explain public polarization and exacerbate societal conflicts. The argument is that inside a context of online fragmentation, opinions are seldom challenged and most often reinforced. Being primarily exposed to corroborating information would lead different parts of the public to drift apart. In fact, Mutz (2002) shows that exposure to dissonant political views enhances political tolerance. Others have argued that exposure to conflicting views is central to a democratic citizenry (Mutz and Martin, 2001; see Habermas, 1996). Furthermore, social psychologists have long proposed that

contact with those who are different decreases prejudice and has long-term positive consequences on social relations (especially the contact hypothesis by Allport, 1954; for an application to online contexts, see Amichai-Hamburger and McKenna, 2006). Hence, we should be concerned about online fragmentation, as it may have detrimental consequences for social relations in society at large.

As mentioned in the 'Introduction' section, political and societal polarization resulting from online fragmentation can have specifically tangible consequences in multi-ethnic societies with a history of armed conflict, where inter-ethnic relations often remain tense. This leads us to the question of whether ethnic groups are as fragmented as groups defined by political ideology. Thelwall (2009) offers evidence of ethnic homophily between *MySpace* users in the United States. However, the United States is a relatively 'tame' context. Post-conflict societies are often characterized by deep mistrust between ethnic groups (cf. Håkansson and Sjöholm, 2007). As the offline world is segregated (e.g. schools, the media and public administration), ethnic segregation may be perceived as something normal in such countries. At the same time, online spaces could, in the absence of physical barriers and a low visibility of ethnic markers, provide interaction opportunities across ethnic divides. Still, ethnic markers such as language or name (e.g. Dunning and Harrison, 2010; Habyarimana et al., 2007) remain often at least partly visible in the online sphere, enabling people to intuitively identify the ethnicity of their counterpart. Given the existence of ethnic markers and as a result from mistrust and habit from offline networks, people should therefore strongly prefer interaction partners from their own ethnic groups even in online contexts. This constitutes our first hypothesis:

Hypothesis 1. Online interactions in post-conflict societies exhibit ethnic segregation.

Once we have established the existence and extent of online segregation in a multi-ethnic society, we are interested in whether active engagement with politics makes individuals choose more homogeneous networks. The existing literature assumes that politics is an important factor explaining ethnic divides in multi-ethnic societies. Proponents of a constructivist understanding of ethnicity assume that ethnicity becomes salient when political actors, so-called 'ethnopolitical entrepreneurs' (Brubaker, 2002), have an interest in mobilizing along ethnic lines (e.g. Bates, 1974; Ferree, 2006; Posner, 2005). According to this argument, Eifert et al. (2010) show that the importance of ethnic identity among average citizens increases with the temporal proximity of presidential elections: the closer the next election, the more likely are respondents to name ethnicity as their most important identity. The authors explain this by political actors 'playing the ethnic card' as a means to gain votes, which in turn impacts how average citizens relate to their ethnic identity. Michelitch (2015) finds that non-coethnics are more disadvantaged in taxi price negotiations in Ghana around election time. More specifically, non-coethnics have to pay higher prices for a taxi ride during election time. Michelitch (2015) explains this by the fact that the election campaign creates an ethnically heated environment that makes individuals discriminate along ethnic lines. We argue that politically engaged individuals should have more segregated personal networks for two reasons. First, politics in multi-ethnic societies often constitutes a contested, emotionally and

ethnically loaded issue. Research on selective exposure and information seeking has shown that individuals have a higher preference for consonant information when they feel threatened (Lavine et al., 2005). Since politics frequently constitutes a contested issue in these contexts, politically engaged individuals may more strongly seek consonant information than others. Yet, individuals will be more likely to find consonant information among their coethnics on the topics they care about if we assume that political opinions are more ethnically homogeneous within than across ethnic groups.² Second, politically engaged individuals may just care more about ethnicity if they live in a context where ethnicity and politics are frequently evoked together: as politically engaged individuals, they are more frequently exposed to the issue of ethnicity. As a result, they may develop a stronger preference for coethnics and discriminate more along ethnic lines when choosing their sources of information. In consequence of these two factors, we expect that politically engaged individuals have more ethnically segregated personal networks. This constitutes our second hypothesis:

Hypothesis 2. Political bloggers have more segregated networks than non-political bloggers.

In the subsequent section, we present our case and the research design that tests these hypotheses.

Case selection and methods

In the following section, we first introduce the case our study focuses on: Bosnia and Herzegovina, and the country's largest blogger platform. We then give an overview of the methods we employ, specifically aspects of the measurement of ethnicity, segregation and political engagement.

Bosnia and Herzegovina

We choose the Bosnian case for different methodological and practical reasons. Bosnia and Herzegovina is representative for a country where ethnic identity and ethnic divisions transcend into public administration and politics. Bosnia's constitution sets out three ethnically defined 'constituent people': ethnic Bosniaks, Serbs and Croats (Bochsler, 2012; Claridge, 2010; Milanovic, 2010). Most political parties are linked to one of the three ethnic groups (Šedo, 2010). Furthermore, political institutions give ethnic groups extensive veto rights that they can and do use against one another (Bahtić-Kunrath, 2011). Political mobilization often takes the shape of appeals to ethnicity (e.g. Hulsey, 2010).

Before the war of the 1990s, the inhabitants of Yugoslavia were considered to speak one common language, Serbo-Croatian. After the war, however, language was used as a means to enhance ethnic identities, and thus, differences between languages were encouraged by political actors, for example, through the publication of new grammar books (Bugarski, 2012; Okuka, 1998). As a result, language is today a marker of ethnic identity

(e.g. Tolimir-Hölzl, 2009, 2011a, 2011b).³ At the same time, differences between languages have remained small (e.g. Kordić, 2008). This signifies that language barriers are negligible and that all ethnic groups can communicate with ease with one another, while being able to infer the ethnicity of their communication partners from their language. For our research, this constitutes a unique situation. The existing language differences allow us to determine the bloggers' ethnicity from the texts they write on their blogs. At the same time, it is important to note that any observed segregation between ethnic groups on the blogger platform cannot be explained by linguistic barriers to communication, as differences between languages have remained very small.

A final consideration for our case selection is the fact that many Bosnians are active Internet users. By the end of 2013, about 58% of the Bosnian population had Internet access (Communications Regulatory Agency, 2014; numbers for later years are not yet available). Furthermore, blogging is a common way in Bosnia of sharing one's thoughts and connecting with old and new friends. While there are other smaller websites used by Bosnian bloggers, we focus on the country's largest blogger platform, *blogger.ba*. We introduce this platform in more detail in the following section.

Blogger platform: Data collection and sample description

As mentioned in the introduction, *blogger.ba* belongs to the top local websites in Bosnia and Herzegovina, and has reached about 20% of the Bosnian population in 2011 (Džihana et al., 2012). According to the website's own information, about 340,000 bloggers write on some 312,000 blogs on this site – statistically, this is one blogger for each seventh Internet user in Bosnia and Herzegovina. All bloggers have a personal profile where they publish sociodemographic information on themselves, such as their age, gender and hometown. Also, importantly for our purpose, blogger profiles include a list of the blogger's favourite blogs, comparable to friendship ties on social networking sites. Since many bloggers make active use of this feature, the blogger platform resembles a large social network.

All data used for this study are publicly available on the website, and no registration is necessary to access the data. For our data collection, we use a Python script that automatically extracts all necessary data. Data were collected between August 2015 and February 2016; the study is based on the state of the blogger network in February 2016. To obtain the sample, we use the following procedure: we start with a randomly chosen blogger and obtain all favourite markings of this blogger. In the next step, we obtain all favourite markings of the newly added bloggers and then repeat the previous step. We stopped the procedure when no more than a few new bloggers were added in each step.⁴ As a result, more active and popular bloggers have a higher likelihood to end up in the sample. This does not constitute a problem, as we are especially interested in bloggers with a certain level of activity and popularity. The procedure results in an original sample of 82,886 bloggers.

In a second step, we restrict the sample to bloggers with a certain level of activity: we only include bloggers who follow at least 10 blogs, and who are followed by at least 10 others. Furthermore, we only include bloggers that have authored at least 10 blog posts. Choosing a lower threshold would leave us with a bigger sample (see section B of the

online appendix for exact details). However, setting the threshold sufficiently high is important for the subsequent analysis, since we are able to determine the bloggers' language (and thereby their ethnicity) only if they have authored a sufficient number of blog posts. Furthermore, we are interested in the level of ethnic segregation inside the bloggers' connections, which is why we include only bloggers with a sufficient number of connections. Finally, restricting the sample in this manner allows us to effectively exclude spam blogs, which are written either with the aim of promoting a certain product or simply contain meaningless text. Spam blogs are only rarely followed by other bloggers and are therefore effectively excluded using this threshold.

This procedure leaves us with 5383 bloggers, 6111 blogs, 667,367 blog posts and 439,625 connections between bloggers. Summary statistics including sociodemographic variables are found in section A of the online appendix.

Methods

In the following section, we present our research design. This includes (1) the procedure for determining the ethnicity of bloggers, (2) our measure of ethnic segregation and (3) determining whether a blogger is politically engaged.

Determining the ethnicity of bloggers. As mentioned earlier, we take advantage of slight language differences between Bosnian, Croatian and Serbian (see Tolimir-Hözl, 2009, 2011a, and our own analysis in section C of the online appendix). More specifically, we build on the approach of computer linguists (Tiedemann and Ljubešić, 2012), who provide lists of words that effectively distinguish between Bosnian, Serbian and Croatian. We use these lists to compute three scores for each blog post that indicate how likely a post is written in each of the three languages (Bosnian, Serbian and Croatian). We sum up the three language scores by blogger and assume that the most likely language of a given blogger is the one with the highest additive score. This procedure takes into account the fact that bloggers might sometimes mix languages, and allows us to average over a large number of texts which were often written over an extended period of time.

Using this procedure, we find that 76% of the bloggers write mainly Bosnian texts, 16% write in Croatian and 7% in Serbian.⁵ Compared to data from the 2013 Bosnian census (Agency of Statistics of Bosnia and Herzegovina, 2016), the share of ethnic Serbs is lower and the share of ethnic Bosniaks is higher among bloggers than among the population (50% of the Bosnian population are ethnic Bosniak, 15% are ethnic Croat and 30% are ethnic Serb). We believe that this asymmetry can be mainly explained by the fact that bloggers with Serb ethnicity, especially those with strong ethnic ties, have 'migrated' to Serbian platforms, for example, *blog.rs*.

Measuring segregation. For measuring ethnic segregation, we follow the approach by Echenique and Fryer (2007), who argue that a measure of segregation should be disaggregated to the level of individuals. To measure segregation in this manner, we look at connections between bloggers, more specifically favourite markings: bloggers can mark other blogs as their favourites. Since the design of the platform does not allow bloggers to share blogs, marking someone's blog as favourite can be interpreted as a sign of

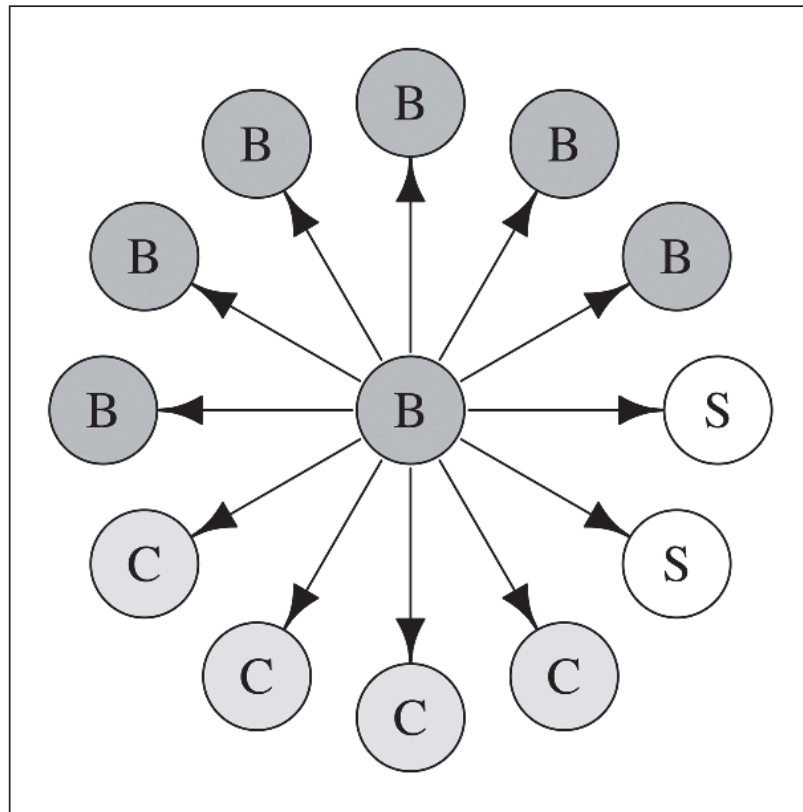


Figure 1. Network of a Bosniak ('B') blogger with other ethnic Bosniak, Serb ('S') and Croat ('C') bloggers; the blogger has an intra-ethnic link share of $6/12 = 50\%$ (graph created using the igraph package for R by Csárdi and Nepusz, 2006).

support or of interest in the blog's owner. In other words, the design of the platform allows us to infer a blogger–blogger connection from the observed blogger–blog connection, and to see whether bloggers who mark each other as favourites tend to be members of the same ethnic group. Note that those markings are directed and not necessarily mutual, that is, if blogger A likes blogger B's blog, blogger B does not necessarily like blogger A's blog. Put differently, favourite markings rather resemble the 'follow' function on Twitter than friendship ties on Facebook.

We quantify the level of a blogger's ethnic segregation by calculating his or her intra-ethnic link share. The intra-ethnic link share designates the number of favourite markings of coethnic bloggers divided by the number of all existing favourite markings of a given blogger. Mathematically, the intra-ethnic link share for a given blogger b is $l_{b,e} / l_b$, where $l_{b,e}$ are the number of links that b has with members of her own ethnic group e , and l_b are all links b has.⁶ Figure 1 illustrates the concept of intra-ethnic link share for a hypothetical Bosniak blogger. For this blogger, the total number of links l_b is 12, and the number of links to bloggers from her own ethnic group $l_{b,e}$ is 6. Hence, this blogger's intra-ethnic link share is $l_{b,e} / l_b = 6 / 12 = 0.5 = 50\%$. Put differently, 50% of this blogger's connections go to bloggers from the same ethnic group.

To test hypothesis 1, we examine whether the intra-ethnic link share in the actually observed network is higher than what we would observe if bloggers connected with randomly selected others. In other words, we compare the observed segregation with a baseline segregation in networks in which ethnicity is randomly assigned, that is, plays no

role (cf. for this approach, Echenique and Fryer, 2007; Hubert and Schultz, 1976). To this end, we simulate a large number of networks ($N = 10,000$) in which we permute or ‘scramble’ bloggers’ ethnicity. This signifies that the shares of Bosniaks, Serbs and Croats in the simulated networks correspond to the respective shares in the observed network, and that all other network features are held constant, such as the number of connections of each bloggers. For each simulated network, we then calculate the intra-ethnic link share per blogger and then average the intra-ethnic link share across all bloggers. If our first hypothesis holds, the average intra-ethnic link share in the actually observed network should be significantly higher than in the simulated networks.

Political blogs and political bloggers. To test hypothesis 2, we need to categorize bloggers into those who are politically engaged and those who are not. The design of the platform requires bloggers to assign their blog to one category from a set list, including categories such as politics, sports, technology and love. To discern whether a blogger is politically engaged or not, we rely on the bloggers’ own choice of category. We define a *political blogger* as a blogger who owns at least one political blog (besides their political blog, political bloggers can run other non-political blogs), and a *political blog* as a blog in the category ‘politics’. Using this categorization, we determine that 2.5% of all bloggers in our sample are politically engaged.

To check whether the bloggers’ own choice of category is a meaningful indicator of their political engagement, we let a native student assistant code 500 randomly drawn blog posts with regard to their political content (we oversampled political blogs and texts from Serb and Croat bloggers). We find that 94% of blog posts from non-political blogs have non-political content. On the other hand, 54% of blog posts from political blogs have indeed political content. This signifies that bloggers identifying themselves as political in fact seem to write frequently about political topics but also have other interests (46% of blog posts from political bloggers have non-political content). Bloggers who identify as non-political, on the other hand, cover only very seldom political topics (in 6% of their blog posts).

This allows us to tackle our second hypothesis, where we examine the difference in the intra-ethnic link share of political and non-political bloggers. To test this hypothesis, we run again simulations. During each simulation, we permute (‘scramble’) ethnicity separately for political and non-political bloggers, which is necessary because the respective shares of the three ethnic groups differ between political and non-political bloggers. We then calculate the level of ethnic segregation for each blogger and calculate separately the averages for political and non-political bloggers. For each simulation, we then calculate the difference between the simulated and the observed intra-ethnic link share, again separately for political and non-political bloggers. If our second hypothesis holds, the difference between actual and simulated segregation should be significantly higher for political than for non-political bloggers. This would signify that political bloggers have ethnically more segregated networks than non-political bloggers.

Results

In the following, we present our findings with regard to hypothesis 1 (overall ethnic segregation in the network) and hypothesis 2 (higher ethnic segregation for political bloggers).

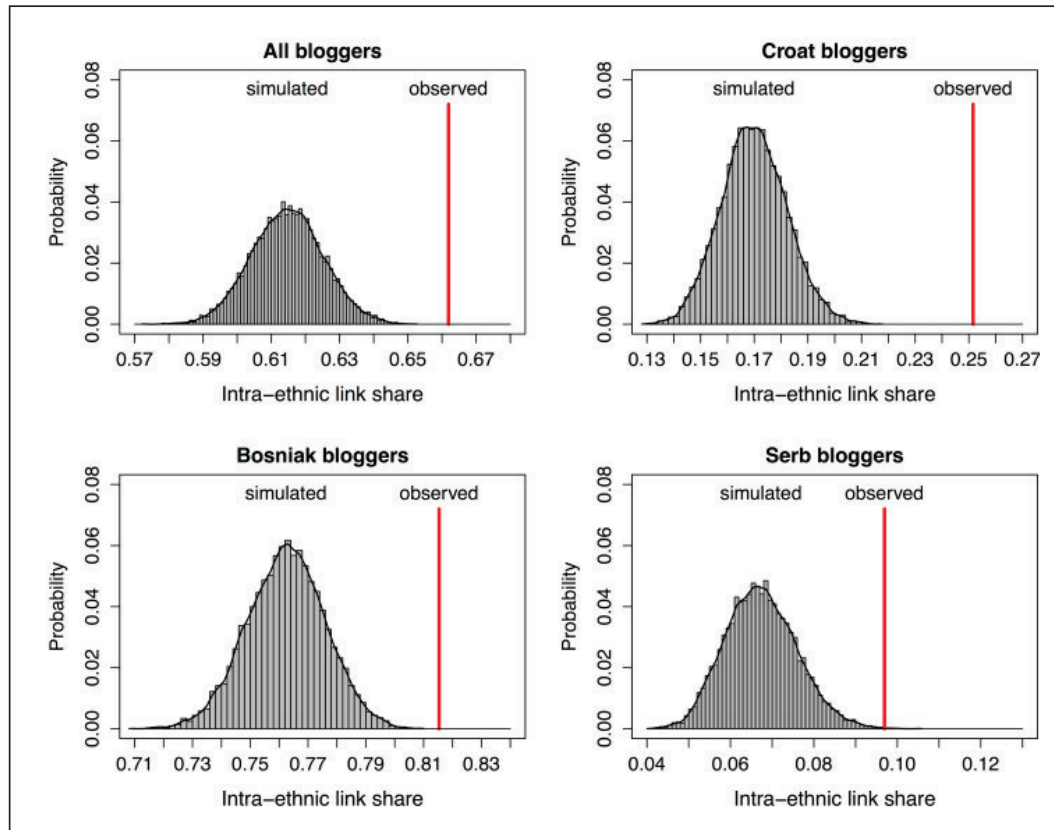


Figure 2. Distribution of the mean intra-ethnic link share in 10,000 random simulated networks for all bloggers, Bosniak, Croat and Serb bloggers.

Evidence of ethnic segregation

To test our first hypothesis, we compare the observed blogger network with 10,000 simulated networks. As outlined in the methods section above, we randomly vary the ethnicity of bloggers, while not changing the existing links between bloggers. Each time we simulate a network, we calculate the intra-ethnic link share per blogger, the average of all intra-ethnic link shares, as well as the average of the intra-ethnic link shares per ethnicity.

The panels in Figure 2 display the results from our simulations. As the upper-left panel of the figure shows, the mean intra-ethnic link share in the observed network is considerably higher than the mean intra-ethnic link share in the simulated networks. The observed mean value over all ethnicities is 66.2%, but only 61.5% in the simulated networks. In other words, whereas 61.5% of all connections between bloggers in the random networks occur between bloggers of the same ethnicity, the number of those connections amounts to 66.2% in the observed network. Expressed in standard deviations (*SDs*), the observed value lies 4.48 *SDs* above its expected value. In terms of randomization inference (Gerber and Green, 2012), none of the simulated networks have an intra-ethnic link share as high as the observed network. This corresponds to a *p* value of less than .0001 (cf. Gerber and Green, 2012), implying that ethnicity plays an important and highly significant role for the establishment of connections between bloggers.

Next, we need to check that findings are not explained by exclusively one ethnic group being highly ethnically segregated, but that the phenomenon holds for each ethnic group in the network. Therefore, we look at the level of ethnic segregation inside each group separately. For Bosniak bloggers, the observed segregation lies 5.17 *SDs* (or 5.3 percentage points) above the expected segregation. For Croats, it is 6.02 *SDs* (8.2 percentage points), and for Serbs, it is 3.9 *SDs* (3.0 percentage points). The observed value of the intra-ethnic link share lies above any of the 10,000 simulated values both for Bosniak and Croat bloggers, implying a *p* value of less than .0001 for both Bosniak and Croat bloggers. For Serb bloggers, 7 out of 10,000 simulations reach an intra-ethnic link share at least as high as the observed value, implying a *p* value of .0007 (cf. Gerber and Green, 2012). This signifies that ethnic segregation is high for each ethnic group and is not dependent on which ethnic group we take into consideration.

Do politically engaged bloggers have more segregated networks?

To test hypothesis 2, we examine the difference in the average intra-ethnic link share between political and non-political bloggers. As outlined above, we expect that political bloggers should be more ethnically segregated than other bloggers, signifying that their intra-ethnic link share should be higher. As mentioned earlier, we use the bloggers' own choice of category to determine whether a blog and blogger is political or not.

75.2% of the links of political bloggers go to bloggers from the same ethnic group, whereas for non-political bloggers, this value is only 66.0%. From this, we cannot directly deduct that political bloggers have a more segregated network than non-political bloggers, as more Bosniaks are political bloggers. At the same time, Bosniaks have by default a higher intra-ethnic link share, as they constitute the majority in the network: even if they choose their connections randomly, they are still more likely to choose someone from their own ethnic group. Yet, we are interested only in that part of ethnic segregation that cannot be explained by the respective shares of the ethnic groups in the sample.

To test whether political bloggers have indeed a more segregated network than non-political bloggers, we therefore run 10,000 simulations where we permute ('scramble') the ethnicities of bloggers. As the different ethnic groups have different likelihoods of becoming political bloggers (in fact, the majority of political bloggers are ethnic Bosniaks), we permute in each simulation ethnicity separately within the group of political bloggers and within the group of non-political bloggers. In other words, political bloggers in the simulated networks have exactly the same likelihood of being Bosniak as political bloggers in the actual network; the same goes for Croats and Serbs.

In contrast to our approach for testing hypothesis 1, in each simulation, we average the intra-ethnic link shares separately for political and non-political bloggers. We again calculate the difference between the observed and simulated value of the intra-ethnic link share of political and non-political bloggers, and compare how much these values differ from the simulated ones. The results are displayed in Figure 3. As we can see in the figure, the actual value of ethnic segregation is significantly higher than the simulated values for both political and non-political bloggers ($p < .05$ for political bloggers and $p < .0001$ for non-political bloggers). This signifies that both political and non-political

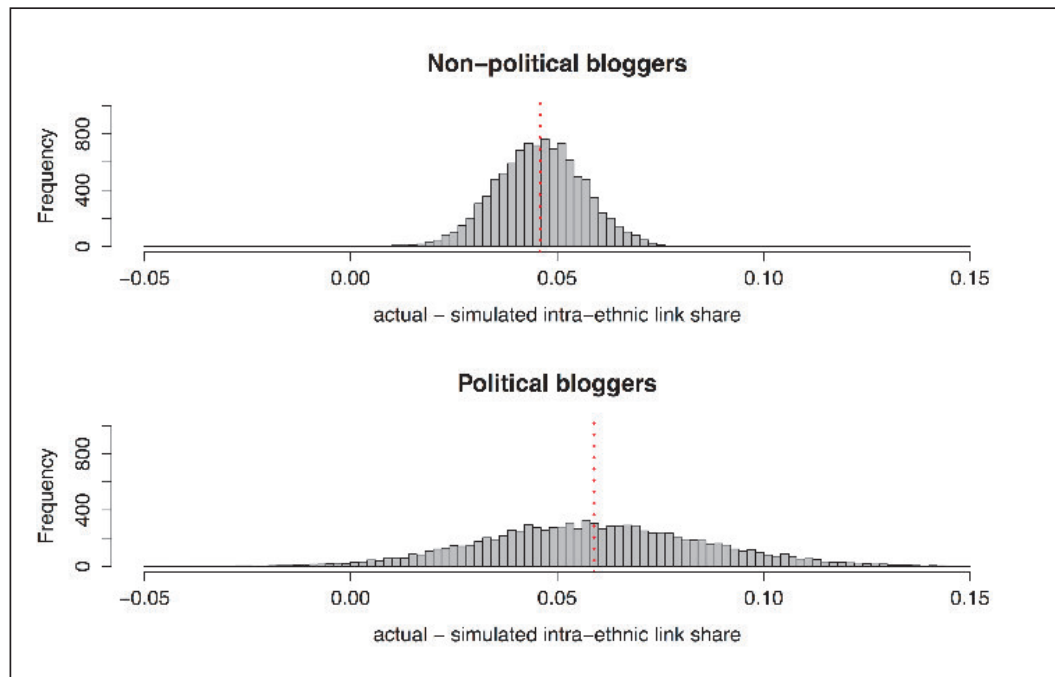


Figure 3. Distributions of the difference between the actually observed and the simulated intra-ethnic link share separately for political and non-political bloggers in 10,000 random simulated networks. The red dashed lines mark the respective means of the distributions.

bloggers are guided by ethnic affiliation when establishing connections among one another. Actual segregation exceeds random segregation on average by 4.6 percentage points ($SD=0.011$) for non-political bloggers and by 5.9 percentage points ($SD=0.027$) for political bloggers. The difference in means is significant in a one-sided t test ($t(12953)=-45.20$; $p<.0001$). This means that political bloggers are indeed significantly more ethnically segregated than non-political bloggers, even when taking into account the different sizes of ethnic groups and a greater likelihood of becoming political bloggers among Bosniaks. In total, this confirms our hypothesis that politically engaged bloggers have ethnically more segregated personal networks than other bloggers.

Conclusion

Recently, it has been argued that the segregation of online networks contributes to a polarization of society at large (e.g. El-Bermawy, 2016; Wong et al., 2016). Our study is the first to examine the existence and extent of online segregation between bloggers from different ethnic groups in a post-conflict, multi-ethnic society. Furthermore, we take a closer look at the role that politics plays in increasing online segregation. More specifically, we investigate whether bloggers writing about political topics have more segregated personal networks than other, less politically engaged and interested bloggers. We expect this to be the case as politics in multi-ethnic, post-conflict societies often constitutes a highly contested and ethnically loaded issue that could increase the concerned bloggers' desire for an ethnically homogeneous environment. We study ethnic segregation on the largest blogger platform of Bosnia and Herzegovina. Using Monte Carlo

simulations, we find strong evidence for ethnic segregation across the platform that cannot be explained by network features or sociodemographic characteristics of the bloggers. Moreover, in line with our expectations, political bloggers have indeed ethnically more segregated personal networks than other bloggers.

Overall, our results suggest at least two important conclusions. First, while geographical boundaries, social control and the visibility of ethnic markers are much lower on social media than in the offline world, bloggers in the multi-ethnic society we study still self-select into mostly ethnically homogeneous environments. Expecting that a broad public exchange transcending ethnic categories will spontaneously develop online will mostly likely remain a vain hope. Second, the results shed light on how a context of ethnicized politics interacts with individuals' behaviour. While we cannot tell whether a preoccupation with politics makes individuals more concerned about ethnicity, or whether those who care more about ethnicity start caring more about politics, we can conclude that those who currently dominate the online political debate exhibit a stronger discriminatory behaviour along ethnic lines. This begs the question of who shapes online political debates. In the multi-ethnic, post-conflict context that we study, at least, it appears that the online political debate is not dominated by those who overcome ethnic categories but by those who in their social interactions put more weight on ethnic categories than the average.

For future studies, it may be useful to examine whether the dynamics inside the blogger network are additionally impacted by political events, such as elections. For the offline context, existing studies have offered evidence that inter-ethnic relations become more tense around election times. Furthermore, it may be useful to examine whether the design, that is, the algorithms of online portals (cf. Bozdag and van den Hoven, 2015), can mitigate the degree of online segregation in general and the exacerbating impact of politics on online ethnic segregation in particular.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The authors gratefully acknowledge funding from the Alexander von Humboldt Foundation (Sofja Kovalevskaja Award for Nils B. Weidmann).

Notes

1. It is, to our best knowledge, the first study examining online ethnic segregation in a post-conflict setting. Ruesch (2011), focusing on a context of active conflict (Israel/Palestine), conducts a qualitative study of Facebook groups and comes to the conclusion that 'virtual spaces bear a potential for increased intergroup communication, yet these potentials are only realized to a very limited degree'. Our study differs in methodology and scale, and in the fact that we take into consideration the influence of political engagement on ethnic segregation.
2. In contexts where a strong alignment between vote choice and ethnic identity is prevalent (see, for example, Huber, 2012), as is the case in many multi-ethnic societies, we have all reason to expect a stronger homogeneity of political opinions within than across ethnic groups.

3. The statistical correlation between ethnicity and language is higher than 99% for all three ethnic groups (see section C of the online appendix).
4. The number of new bloggers per day dropped from 20,816 at the beginning of the process to 425 on the last day of the collection effort.
5. The language of 0.8% of the bloggers could not be coded with certainty, which is why we exclude them from the sample.
6. If a blogger has links to several blogs of the same blogger, we count each of those links.

ORCID iD

Nils B Weidmann  <https://orcid.org/0000-0002-4791-4913>

References

- Adamic LA and Glance N (2005) The political blogosphere and the 2004 U.S. election: Divided they blog. In: *Proceedings of the 3rd international workshop on link discovery*, Chicago, IL, 21–25 August, pp. 36–43. New York: ACM.
- Agency of Statistics of Bosnia and Herzegovina (2016) *Results of the 2013 census of population, households and dwellings in Bosnia and Herzegovina*. Available at: <http://www.popis2013.ba> (accessed 21 September 2016).
- Allport GW (1954) *The Nature of Prejudice*. Reading, MA: Addison-Wesley.
- Amichai-Hamburger Y and McKenna KYA (2006) The contact hypothesis reconsidered: Interacting via the Internet. *Journal of Computer-Mediated Communication* 11(3): 825–843.
- Bahtić-Kunrath B (2011) Of veto players and entity-voting: Institutional gridlock in the Bosnian reform process. *Nationalities Papers* 39(6): 899–923.
- Barberá P, Jost JT, Nagler J, et al. (2015) Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science* 26(10): 1531–1542.
- Bates R (1974) Ethnic competition and modernization in contemporary Africa. *Comparative Political Studies* 6(4): 457–484.
- Bochsler D (2012) Non-discriminatory rules and ethnic representation: The election of the Bosnian State Presidency. *Ethnopolitics* 11(1): 66–84.
- Bozdag E and van den Hoven J (2015) Breaking the filter bubble: Democracy and design. *Ethics and Information Technology* 17(4): 249–265.
- Brubaker R (2002) Ethnicity without groups. *European Journal of Sociology* 43(2): 163–189.
- Bugarski R (2012) Language, identity and borders in the former Serbo-Croatian area. *Journal of Multilingual and Multicultural Development* 33(3): 219–235.
- Chandra K (2006) What is ethnic identity and does it matter? *Annual Review of Political Science* 9: 397–424.
- Claridge L (2010) *Discrimination and Political Participation in Bosnia and Herzegovina, Sejdic and Finici v. Bosnia and Herzegovina*. London: Minority Rights Group International.
- Colleoni E, Rozza A and Arvidsson A (2014) Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using Big Data. *Journal of Communication* 64(2): 317–332.
- Communications Regulatory Agency (2014) Annual report of the Communications Regulatory Agency. Available at: <http://rak.ba/eng/index.php?uid=1273696422> (accessed 29 October 2014).
- Csárdi G and Nepusz T (2006) The igraph software package for complex network research. *International Journal of Complex Systems* 1695(5): 1–9.

- Dunning T and Harrison L (2010) Cross-cutting cleavages and ethnic voting: An experimental study of cousinage in Mali. *American Political Science Review* 104(1): 21–39.
- Džihana A, Ćendić K and Tahmaz M (2012) Mapping digital media: Bosnia and Herzegovina. *Open Society Foundations Report*. Available at: <http://www.opensocietyfoundations.org/sites/default/files/mapping-digital-media-bosnia-20120706.pdf> (accessed 25 April 2014).
- Ebo BL (ed.) (1998) *Cyberghetto or Cybertopia? Race, Class, and Gender on the Internet*. Westport, CT: Praeger.
- Echenique F and Fryer RG (2007) A measure of segregation based on social interactions. *Quarterly Journal of Economics* 122: 441–485.
- Eifert B, Miguel E and Posner DN (2010) Political competition and ethnic identification in Africa. *American Journal of Political Science* 54(2): 494–510.
- El-Bermawy MM (2016) Your filter bubble is destroying democracy. *Wired*, 18 November. Available at: <https://www.wired.com/2016/11/filter-bubble-destroying-democracy/> (accessed 19 December 2016).
- Etling B, Kelly J, Faris R, et al. (2010) Mapping the Arabic blogosphere: Politics and dissent online. *New Media & Society* 12(8): 1225–1243.
- Ferree KE (2006) Explaining South Africa's racial census. *Journal of Politics* 68(4): 803–815.
- Garrett RK (2009) Echo chambers online? Politically motivated selective exposure among Internet news users. *Journal of Computer-Mediated Communication* 14(2): 265–285.
- Gerber AS and Green DP (2012) *Field Experiments: Design, Analysis, and Interpretation*. New York: W. W. Norton & Company.
- Gitlin T (1998) Public sphere or public sphericules? In: Liebes T and Curran J (eds) *Media, Ritual and Identity*. London: Routledge, pp. 168–174.
- Habermas J (1996) *Die Einbeziehung des Anderen: Studien zur Politischen Theorie* [The Inclusion of the Other: Studies on Political Theory]. Frankfurt am Main: Suhrkamp Verlag.
- Habyarimana J, Humphreys M, and Posner DN, et al. (2007) Why does ethnic diversity undermine public goods provision? *American Political Science Review* 101(4): 709–725.
- Håkansson P and Sjöholm F (2007) Who do you trust? Ethnicity and trust in Bosnia and Herzegovina. *Europe-Asia Studies* 59(6): 961–976.
- Huber JD (2012) Measuring ethnic voting: Do proportional electoral laws politicize ethnicity? *American Journal of Political Science* 56(4): 986–1001.
- Hubert L and Schultz J (1976) Quadratic assignment as a general data analysis strategy. *British Journal of Mathematical and Statistical Psychology* 29(2): 190–241.
- Hulsey JW (2010) 'Why did they vote for those guys again?' Challenges and contradictions in the promotion of political moderation in post-war Bosnia and Herzegovina. *Democratization* 17(6): 1132–1152.
- Kelly J and Etling B (2008) Mapping Iran's online public: Politics and culture in the Persian blogosphere. *Berkman Center for Internet and Society and Internet & Democracy Project, Harvard Law School*. Available at: http://cyber.harvard.edu/sites/cyber.harvard.edu/files/Kelly&Etling_Mapping_Irans_Online_Public_2008.pdf (accessed 11 January 2017).
- Kordić S (2008) Nationale Varietäten der Serbokroatischen Sprache [National Varieties of the Serbo-Croatian Language]. In: Golubović B and Raecke J (eds) *Bosnisch, Kroatisch, Serbisch als Fremdsprachen an den Universitäten der Welt* [Bosnian, Croatian and Serbian as Foreign Languages at Universities Around the World]. München: Kubon & Sagner, pp. 93–102.
- Lavine H, Lodge M and Freitas K (2005) Threat, authoritarianism, and selective exposure to information. *Political Psychology* 26(2): 219–244.
- Michelitch K (2015) Does electoral competition exacerbate interethnic or interpartisan economic discrimination? Evidence from a field experiment in market price bargaining. *American Political Science Review* 109(1): 43–61.

- Milanovic M (2010) Sejdic and Finci v. Bosnia and Herzegovina. *American Journal of International Law* 104: 636–640.
- Morozov E (2012) *The Net Delusion: The Dark Side of Internet Freedom*. New York: PublicAffairs.
- Mutz DC (2002) Cross-cutting social networks: Testing democratic theory in practice. *American Political Science Review* 96: 111–126.
- Mutz DC and Martin PS (2001) Facilitating communication across lines of political difference: The role of mass media. *American Political Science Review* 95(1): 97–114.
- Okuka M (1998) *Eine Sprache, viele Erben: Sprachpolitik als Nationalisierungsinstrument in Ex-Jugoslawien* [One Language, Multiple Heirs: Language Politics as a Means for Nationalization in the Former Yugoslavia]. Klagenfurt: Wieser Verlag.
- Pariser E (2011) *The Filter Bubble: What the Internet Is Hiding from You*. London: Penguin.
- Posner DN (2005) *Institutions and Ethnic Politics in Africa*. New York: Cambridge University Press.
- Prior M (2007) *Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections*. Cambridge: Cambridge University Press.
- Ruesch M (2011) A peaceful net? Intergroup contact and communicative conflict resolution of the Israel-Palestine conflict on Facebook. In: *First global conference on communication and conflict*, Prague. Available at: <http://www.lse.ac.uk/media@lse/study/pdf/Michelle-Ruesch-CCConference-forLSE-final.pdf> (accessed 10 January 2017).
- Šedo J (2010) The party system of Bosnia and Herzegovina. In: Stojarová V and Emerson P (eds) *Party Politics in the Western Balkans*. London; New York: Routledge, pp. 85–98.
- Sunstein CR (2001) *Republic.com 2.0*. Princeton, NJ: Princeton University Press.
- Thelwall M (2009) Homophily in MySpace. *Journal of the American Society for Information Science and Technology* 60(2): 219–231.
- Tiedemann J and Ljubešić N (2012) Efficient discrimination between closely related languages. In: Kay M and Boitet C (eds) *Proceedings of the 2012 COLING (International Conference on Computational Linguistics) conference*. Mumbai, India, pp. 2619–2634.
- Tolimir-Hölzl N (2009) *Bosnien und Herzegowina: Sprachliche Divergenz auf dem Prüfstand* [Bosnia and Herzegovina: Evaluating Linguistic Divergence]. München: Kubon & Sagner.
- Tolimir-Hölzl N (2011a) ‘Die Unterschiede sind nun da’: Zur sprachlichen Entwicklung in Bosnien und Herzegowina [‘The Differences Have Come to Exist’: On Linguistic Evolution in Bosnia and Herzegovina]. *Zeitschrift für Slawistik* 56(3): 319–335.
- Tolimir-Hölzl N (2011b) Linguistic divergence in Bosnia: Considerations about vertical and horizontal leveling. *International Journal of the Sociology of Language* 2011(212): 55–68.
- Tremayne M, Zheng N, and Lee JK, et al. (2006) Issue publics on the web: Applying network theory to the war blogosphere. *Journal of Computer-Mediated Communication* 12(1): 290–310.
- Vogt M (2014) The Ethnic Power Relations (EPR) core dataset 2014: Codebook. Available at: http://www.icr.ethz.ch/data/epr/EPR-2014_Codebook.pdf.
- Wong JC, Levin S and Solon O (2016) Bursting the Facebook bubble: We asked voters on the left and right to swap feeds. *The Guardian*, 16 November. Available at: <https://www.theguardian.com/us-news/2016/nov/16/facebook-bias-bubble-us-election-conservative-liberal-news-feed> (accessed 20 March 2017).