Boon or Bane for Political Engagement:
A Large-Scale Study of Normalization of Facebook

Abstract

Does Facebook help or hurt political engagement? Some scholars argue (e.g., Resnick, 1997, 1998) that social media has taken over old media (e.g., newspaper, television) as a leading communication channel, but communication practices of human beings remain unchanged on the web. Resnick suggests the internet neither empowers citizens nor revitalizes democracy, because online communications occur within preexisting economic, legal, and social frameworks. To test this theory called normalization of the cyberspace, this study examines engagement on two verified Facebook pages run by the campaigns of two main candidates of the 2016 U.S. presidential election—Donald Trump and Hillary Clinton. The study combines two computational methods—sentiment analysis and automated topic classification—and a quantitative content analysis method to analyze 6,122 status messages, 2.8 million user comments, and 0.6 million replies. Results largely support the claims of the theory that traditional communication practices of election campaigns are transferred to the internet.

Keywords: big data, election campaign, machine learning, normalization, political communication, social media

Paper submitted to ICA Computational Methods Interest Group for presentation at the 2018 annual conference in Prague, Czech Republic
Boon or Bane for Political Engagement:
A Large-Scale Study of Normalization of Facebook

As more researchers delve into social media to understand its role in election campaigns, disagreement grows over its contribution to increasing public engagement in political affairs (Boxell, Gentzkow & Shapiro, 2017; El-Bermawy, 2016; Meraz, 2009). Although many experts described social media as a positive force for improving democracy (Chadwick, 2014; Hanson et al., 2010; Kushin & Yamamoto, 2010), evidence pours in to counter this argument. Voter turnout in the 2016 U.S. presidential election was the lowest in 20 years (Wallace, 2016, November 30). Social media stands accused of spreading fake news that many experts and scholars argue skewed the results of the election in an unprecedented way (Allcott & Gentzkow, 2017; Shane, 2017). Moreover, some studies suggested that the networked and interactive nature of social media played no significant role in democratizing the online media ecosystem—neither in diminishing elites' control over media content nor in reining in partisan polarization (Boxell, Gentzkow & Shapiro, 2017). Theoretically, these findings reflect a phenomenon described as the normalization of the Internet (Resnick, 1998), which means social media has taken over old media as a leading communication channel, but communication practices of human beings remain unchanged.

This study tests three concepts supporting the theory of the normalization (Resnick, 1998) by analyzing a large corpus of Facebook posts and user engagement data collected from the campaign pages of the 2016 U.S. presidential candidates Donald Trump and Hillary Clinton. Resnick (1998) proposed that life and politics online would eventually come to resemble life and politics offline. The three related concepts examined in this study are: relational normalization, discursive normalization and selective exposure. The first two concepts focus mainly on the
practices of political elites while the third one focuses on ordinary citizens. *Relational normalization* refers to continuation of unequal power relations between major and minor political actors in cyberspace, in which major actors enjoy more influence and resources than minor actors. *Discursive normalization* refers to the transfer of typical communication practices (e.g., using media to attack opponents) to online. Selective exposure suggests that individuals, regardless of the communication channels they use, expose themselves to information and networks that reiterate their preexisting beliefs (Adamic & Glance, 2005; Hargittai, Gallo, & Kane, 2008). Taken together, these three concepts can help predict behavior of candidates and voters on social media. Evidence supporting these arguments would mean candidates continue to use old campaign techniques (e.g., personal manifestation rather than interaction) while voters join networks of like-minded people on social media.

This interdisciplinary study, conducted by a team of academic researchers from computer science and mass communication departments, combined several computational methods (i.e., a machine learning model of sentiment analysis, and automated topic classification) and a quantitative content analysis to analyze a large corpus of Facebook data. The dataset contains 3.4 million comments and replies on 6,122 Facebook messages posted by the campaigns of Donald Trump and Hillary Clinton in one year from November 8, 2015 to November 7, 2016. The study also analyzed the data relating to other forms of engagement such as reactions (e.g., likes, loves) and shares. Studying Facebook pages of candidates is important because such a study does not only explain how politicians use online tools for campaigns, but also provides a deep understanding of communication practices on online platforms through which candidates communicate with a large number of voters (Foot & Schneider, 2006; Nielsen, 2011; Nielsen & Vaccari, 2013).
The authors of this study developed an in-house software using the Unix system commands and Facebook Graph API to automatically gather posts, comments and replies to comments. Analyzing, understanding and making sense of this data corpus required processing natural language and intelligent computational models such as supervised machine learning based sentiment analyzer and stance detector. The authors developed an automated topic classifier of Facebook posts and used a supervised method for sentiment detection. Little research attempted to build or extend theories of election campaigns on social media through analysis of big data using machine learning models.

This study contributes to political communication and computer science literature in several ways. It analyzed a large dataset to test an emerging theory of political communication. It combines computational and typical content analysis methods to have a deep understanding of election campaign practices on Facebook. It developed a dictionary-based topic detection model to classify policy issues in status messages. It also provides a thorough picture of Facebook users’ engagement on politicians’ pages by analyzing post messages, comments, shares and other forms of engagement (e.g., like, love, angry) in several ways.

**Literature Review**

A review of the literature on the uses of social media in election campaigns indicates a strong role being played by social media in reaching out to voters and engaging them (Davis et al., 2010; Holmes & Sulistiyanto, 2016). Some studies suggested that social networking sites (e.g., Facebook) might help strengthen civic engagement by offering increasing connectivity, interactivity, and cross-ideological exposure (e.g., Chadwick, 2014; Hanson et al., 2010; Kushin & Yamamoto, 2010). But others argued that social media was helping to reach out to those who had already been engaged offline. There was not enough evidence to suggest that social media
was able to engage voters who had previously been nonchalant to political matters (Bennett & Iyengar, 2008; Davis et al., 2010; Macnamara, Sakinofsky, & Beattie, 2012; Nielsen & Vaccari, 2013). A logical conclusion that supports both sides of this argument is that social media has displaced old media as the leading communication channel but it does not have much impact on the communication behaviors of people. Scholars who studied media competition and coexistence (e.g., Dimmick, 2003) would support this conclusion. But from the perspective of political communication, David Resnick's (1998) theory of the normalization of the Cyberspace (p. 49) appears to address this paradox well and deserves attention of researchers for testing and extending the theory.

**Theoretical Framework: Normalization of the Cyberspace**

Resnick (1998) proposed that life and politics on the Internet would come to resemble life and politics in pre-Internet era. He described the hope of "citizen empowerment and the revitalization of democracy" by the Internet (Resnick, 1997, abstract) as unrealistic. Resnick suggested that the Internet would be more crucial than traditional mass media (e.g., newspapers, radio and television) in spreading diverse opinions and controversies because it offers an unprecedented level of connectivity and interactivity. It would serve well the people who are already interested in politics. It would be an effective communication tool for activists to mobilize people, for politicians to campaign, and for businesses to lobby. But it won't bring any major transformation because communications would occur within the old political, economic, social and legal frameworks (Resnick, 1997). Shortcomings of typical communication practices would be transferred to social media where power relations between major (e.g., candidates, parties, institutions) and minor (e.g., individual citizens) political actors would remain unaltered (Schweitzer, 2011). According to Resnick, "Politics on the Web are structured in a double sense,
presenting a structured experience, and reflecting the organized structure of pluralistic political life in the real world. It is truly a creature of modern democratic politics” (p. 49). Schweitzer (2011) identified several concepts relating to the theory of the normalization of cyberspace. Two of those concepts are relevant to this study: relational normalization and discursive normalization.

Relational normalization is defined as continuation of unequal power relations between major and minor political actors in cyberspace, in which major actors enjoy more influence and resources than minor actors (Schweitzer, 2011). Among other consequences of this power relation can be that it reinforces the traditional one-way communication model, in which messages flow from elites to media to citizens—not the other way in most cases (Chadha & Guha, 2016; Williams & Gulati, 2007). This suggests that direct interaction between political elites and ordinary citizens on social media might be rare.

Discursive normalization refers to the transfer of typical communication practices to online. Schweitzer (2011) identified three indicators of discursive normalization in terms of candidates’ communication with voters. Candidates typically use three techniques: (1) strategic news, (2) personalization, and (3) negative campaigning to construct an image, rather than talking about substantial policy issues. Strategic news refers to content that focuses on relative strength of a candidate compared to her or his opponents. Such content includes poll results, advertisements, news and information about media relations and endorsements etc. Candidates prefer sharing strategic news to discussing policy issues (Schweitzer, 2011). Personalization focuses on personal aspects of a candidate such as things happening in personal life, preferences, and emotion etc. Negative campaigning refers to inclination to attack opponents. Studies show these techniques are used in offline campaigns (Kaid & Holtz-Bacha, 2006; Kaid and Strömbäck,
2008). Benoit (1999) identified three similar techniques—self-praise, attacks and defense—to construct image. Studies on media strategies by election campaigns show that communication channels are used more often to transmit information about candidates and campaign issues, generate resources and recruit volunteers than to discuss policy issues and build communities (e.g., Cogburn & Espinoza-Vasquez, 2011; Hooghe & Vissers, 2008; Kluver, 2004; Norris, 2001).

Although Resnick came up with his theory in the early age of the Internet, his propositions hold true in 2017 on social media. The theory can be applied to various other aspects. One of the most compelling but least explored aspects of normalization is patterns of communication among citizens themselves. Understanding this is vital for understanding the future of civic engagement on social media. Social networking sites provide users with a great deal of control over who they connect with (boyd & Ellison, 2007). Studies have shown that users typically connect and spend more time with those who are already known to them offline (Ellison et al., 2010). This has been known as selective exposure, which suggests that individuals expose themselves to information and people who support their preexisting beliefs (Lazarsfeld, Berelson, & Gaudet, 1944). Fan pages of political parties and candidates have been found to attract like-minded people, and discussions on these pages are highly partisan and polarized (Woolley, et al., 2010). boyd and Ellison (2007) wrote, “On many of the large SNS [social network sites], participants are not necessarily networking' or looking to meet new people; instead, they are primarily communicating with people who are already a part of their extended social network” (p. 211). The concept of selective exposure reinforces Resnick’s idea of normalization. It complements the theory by focusing on how users behave on social media.
Political Engagement Online: Indications of Normalization

Social media offers unprecedented opportunities for political engagement (Bennett & Iyengar, 2008; Duggan & Smith, 2016; Harvey, 2014), yet it offers more opportunities to avoid politics by providing more entertainment and non-political content (Prior, 2007). An unprecedented level of connectivity and abundance of information on the Internet have empowered users and created a “pull environment”, in which users control what information they consume and who they talk to (Nielsen & Vaccari, 2013, p. 2337). But there is no concrete evidence to suggest that increasing connectivity and abundant information contribute to increasing civic engagement or better election outcomes. Many young people who are heavy users of social media abandoned political news and embraced entertainment (Ancu, 2015; Duggan, 2015; Mindich, 2005). Da Silva (2013) noted, “The use of Internet for political purposes is minor when compared to other activities” (pp. 177-178). Hindman (2008) stated that about 3% of web traffic is directed to news and media sites, and less than 0.1% is directed to political sites (i.e. websites of politicians). Following a study on online campaigns by 224 candidates in the U.S. Congress elections of 2010, Nielsen and Vaccari (2013) concluded:

Most politicians will not be able to rely on their websites and social media presence to engage in direct communication with the electorate on any significant scale—not so much because politicians do not want to, but because most people do not care much about candidates the way most campaigns are waged in contemporary America. (p. 2351)

Research has shown that most people who engaged in political discussions on social media were already interested in politics (Da Silva, 2013; Zhang, Johnson, Seltzer & Bichard, 2010). Duggan and Smith (2016) found that Americans who had been highly engaged in politics (i.e., ones who vote regularly in elections, volunteer or contribute money to party or campaigns)
were more likely to share political views on social media than those who had been less engaged in politics. Almost 20% of highly engaged users, compared to 6% of the less engaged users, posted comment or message regarding political issues. More than half (53%) of the highly engaged users followed candidates on social media, while 21% less engaged users did so (Duggan & Smith, 2016). However, Pennington, Winfrey, Warner and Kearney (2015) found that following a candidate on Facebook alone does not lead to political engagement.

Some scholars studying effectiveness of social media on civic engagement suggested that the lack of engagement by uninterested citizens may result from selective nature of the Internet (Bimber & Davis, 2003; Ellison et al., 2010). The selective nature, aided by algorithm that helps users find what interests them, poses a major challenge to delivering political news and information to uninterested citizens. Multiple studies have shown that some uninterested citizens may develop interests in politics and eventually start participating in political discussions (Boulianne, 2009; Shah, Cho, Eveland, & Kwak, 2005; Vaccari et al., 2015). But for this to happen, those people will need to be exposed to political news and information on a regular basis. The selective nature also affects those who are more politically engaged and active on social media in a negative way. Studies show political pages on social media are partisan and polarizing in which discussions often take place among like-minded people (Woolley et al., 2010). Users who are more supportive and invested are more likely to be engaged in discussion and support a candidate (Robertson, Vatrapu & Medina, 2010). Supporters of a candidate are less likely than others to hold the candidate accountable (Da Silva, 2013).

In sum, the literature discussed above suggests that civic engagement on the web follows a pattern somewhat similar to real life engagement. The extent to which social media can extend
civic engagement is unclear. The people who engage in political discussions often interact with those who share similar views.

**Use of Social Media in Election Campaigns**

Many politicians running for public office use social media to communicate directly with voters. During his election campaign in 2016, U.S. President Donald Trump wrote, “How do you fight millions of dollars of fraudulent commercials pushing for crooked politicians? I will be using Facebook & Twitter. Watch!” (Trump, 2016). Candidates use social media to construct online presence through personal manifestation (Bimber, 2014; Davis et al., 2010). Hurcombe (2016) suggested that candidates use techniques similar to commercial organizations that seek to build and maintain image on social media. Bimber (2014) who examined former U.S. President Barack Obama’s social media pages during 2008 and 2012 elections identified two techniques used: personalized communication and commodification of social media. Harfoush (2009) showed that Obama constructed his own brand using social media.

Literature suggests that communication on social networking sites is mainly one-way in which candidates disseminate information to voters in order to persuade (Lin, 2015; Sweetser & Lariscy, 2008). In 2006, Facebook gave each congress and gubernatorial candidate in the U.S. midterm elections a page to connect and interact with voters (Williams & Gulati, 2007). But most candidates used these pages to disseminate information about themselves. They responded rarely to any message posted by voters (Sweetser & Lariscy, 2008). Studies on political communication on the web show that political websites espouse the traditional top-down model of communication (e.g., Carlson & Strandberg, 2008; Schweitzer, 2011). For instance, candidates often use their websites to post information about themselves and press release. Interactive features such as commenting, chatting, and wikis are not preferred. Nielsen and
Vaccari (2013) suggested that attention to candidates follows a power-law distribution online, in which a few candidates draw a large number of followers while a majority of candidates gets little attention. Nielsen and Vaccari also suggested that “large-scale direct online communication between politicians and ordinary people via these platforms is a rare” (p. 2333). Lin (2015) found that candidates employ campaign staff in social networks to influence voters’ perceptions of candidates. Because of the selective nature of social media, a candidate may not be able to reach an audience that opposes her or his candidacy. As for President Obama whose social media campaign during 2008 election drew attention of political scientists, “only a very special clientele is reached” through social networking sites (Davis et al., 2010, p. 23). Studies have shown that digital media enable politicians to “preaching to the converted” (Norris, 2003, paper title).

In sum, the literature indicates that behavior of both political elites and ordinary citizens on social media reflect communication behaviors offline. But there is little research that directly examined this phenomenon with a large dataset analyzed through a combination of computational methods and a quantitative content analysis. Since there are not many studies that directly studied the aspects of normalization discussed above, it will not be rational to draw any hypothesis. Therefore, this study asks the following research questions.

**Research Questions:**

RQ1: To what extent, did the 2016 U.S. presidential candidates interact directly with followers on Facebook?

RQ2: To what extent, did the 2016 U.S. presidential candidates discuss policy issues in their Facebook status messages?
RQ3: To what extent, did the 2016 U.S. presidential candidates use Facebook for building image?

RQ4: To what extent, do followers approve or disapprove Facebook status messages of the 2016 U.S. presidential candidates?

RQ5: Which topics discussed in Facebook status messages of the 2016 U.S. presidential candidates attracted more engagement?

Method

User engagement on Facebook comes in various forms (e.g., posting messages, sharing, commenting and reacting) generating a large amount of data of various types such as text, image, video, and statistical data on engagement. A systematic analysis of such a sheer volume of data poses many computational challenges. Some of them are aligned with the challenges in big data analytics and others are specific to election campaigns. This section briefly explains some of the notable challenges regarding collection of data and measures taken by the authors to deal with those challenges. It then moves on to discuss operationalization of concepts and analytical tools used.

Dataset. This study analyzed two Facebook pages run by the campaigns of two main candidates of the 2016 U.S. presidential election—Donald Trump (https://www.facebook.com/DonaldTrump) and Hillary Clinton (https://www.facebook.com/hillaryclinton). The dataset includes a total of 6,122 messages (see distribution in Table 1), 3.4 million comments and replies, and all statistical data on other forms of engagement (e.g., reactions) in one year between November 8, 2015, and November 7, 2016. All messages including text-only messages, videos, photos, and links to webpages were collected. Although the total number of posts published during this period was relatively small
(about 17 per day), the number of comments and replies was large (100,000 comments per day—65 million total). Therefore, only top 500 comments on each post, determined by the number of replies and likes to each comment, were collected. To find out if there was any direct interaction between candidates and followers, top 25 replies to top 100 comments on each post were collected. More data could have yielded better results, but the authors decided to limit the number of comments and replies collected because of time constraints. Reaction (Like, Love, Haha, Wow, Sad, Angry) statistics for each post was also collected.

<table>
<thead>
<tr>
<th>Post Type</th>
<th>Donald J. Trump</th>
<th>Hillary Clinton</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Link</td>
<td>610</td>
<td>1086</td>
<td>1696</td>
</tr>
<tr>
<td>Video</td>
<td>553</td>
<td>986</td>
<td>1539</td>
</tr>
<tr>
<td>Photo</td>
<td>1247</td>
<td>914</td>
<td>2161</td>
</tr>
<tr>
<td>Event</td>
<td>1</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Text only</td>
<td>623</td>
<td>93</td>
<td>716</td>
</tr>
<tr>
<td>Note</td>
<td>0</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>3034</td>
<td>3088</td>
<td>6122</td>
</tr>
</tbody>
</table>

The high volume of data, velocity of generation and the accompanying variety required advanced computational tools that could collect a big dataset, clean, analyze and find meaningful patterns out of it. Laney (2001) defined these challenges as 3Vs of big data. This study used a non-relational data processing tool called MongoDB (Kanade, Gopal, & Kanade, 2014), a high-performance data analysis tool (Pandas) and a real-time data collector (in-house software) to download, manage and analyze the data.

**Accessing Facebook Data.** A second major computational challenge was the difficulty accessing Facebook data. Facebook has a strict policy that makes it hard for researchers to access
its data and, therefore, limits the number of large-scale studies (Alashri et al., 2016) about
election campaigns on Facebook. To overcome this challenge, the authors developed a live
Facebook data collection system using a combination of Unix system commands and Facebook
Graph API.

**Information Extraction and Knowledge Mining.** It was beyond the ability of the
research team to read thousands of status messages and millions of comments and replies to
answer the research questions. So, the authors developed a supervised machine learning method
by leveraging some existing methods to analyze automatically a large amount of social media
content (Hassan, Li, & Tremayne, 2015; Kleinnijenhuis, & De Nooy, 2013). These studies,
however, failed to provide a model of information extraction and knowledge mining, which was
fully suitable for this study. Some models they developed were designed for web content, but
were later used to analyze content of microblogs (e.g., Twitter). To address this problem, the
authors coded a subset of the dataset manually and semi-automatically, and developed a
knowledge mining model to analyze the whole dataset.

**Operationalization of Concepts**

This study examined three aspects of normalization: relational normalization, discursive
normalization, and selective exposure. Relational normalization was operationalized by direct
interaction of candidates with followers on Facebook. Direct interaction was measured by two
things: (1) total number of comments posted by a candidate, and (2) total number of replies a
candidate left on follower comments. The higher the number of comments and replies by
candidates or response to comments, the greater the direct interaction.

Discursive normalization was measured in two ways: (1) automatic detection of topic in
status messages related to policy and non-policy issues; and (2) manual content analysis. The
research team developed an automated topic detection model that included 14 top voting issues identified by Pew Research Center (Top voting issues...2016; More in Topic Detection section).

A quantitative content analysis was also conducted to dig deep into the messages and identify techniques used by candidates to build images. An inter-coder reliability test was done on 5% of status messages (Cohen’s kappa=0.81). Each message was coded using primarily two categories: (1) policy issues and (2) non-policy issues. A message was coded for ‘policy issues’ if it clearly mentions a candidate’s view on any specific policy. One example of such message is: “How Trump would stimulate the U.S. economy”. A message was coded as non-policy issue if it didn’t provide a candidate's view on any specific policy explicitly. The messages coded as ‘non-policy issues’ were further coded for one of three categories: (1) strategic, (2) personalized, and (3) negative campaigning. Schweitzer (2011) suggested that these are three indicators of discursive normalization. A ‘strategic’ post seeks to show the candidate as stronger than opponents. Posts that contain links and/or information about strategic issues (e.g., poll standings; advertisements, endorsement, rallies, media relations, tours) were considered strategic. A personalized post focuses on candidate’s persona (e.g., thankful, emotional, caring) or personal beliefs and ideology, rather than talking directly about the election. A post was considered negative campaigning if the post expressed dislikes for, and/or criticized a person, community, or an institution.

Selective exposure was measured in three ways: (1) overall sentiment of comments, (2) number of unique users who commented on both pages, and (3) reaction to posts in the forms of like, love, haha, sad, wow and angry. A sentiment analysis was conducted to understand the extent to which overall sentiment of comments on a page is: (a) positive towards a candidate, (b) negative towards a candidate and (c) neutral. A highly positive sentiment towards a candidate
would indicate selective exposure by followers on Facebook. A second way to measure is reaction to posts. More positive reactions to all posts in general could indicate selective exposure. Analysis of reactions to posts complemented the sentiment analysis in examining selective exposure. To find out proof of selective exposure, the researchers also looked for unique users who commented on both pages.

**Topic Detection**

A critical component of this study was to identify the topic of a Facebook post. Based on a survey conducted by Pew Research Center (Top voting issues…2016), the research team populated a list of 14 most important topics. The list includes--economy, terrorism/national security, foreign policy, healthcare, gun policy, immigration, social security, veteran affairs, education, Supreme Court appointments, equality, trade policy, environment, abortion, and treatment of LGBT. For each of these topics, the researchers manually curated a list of relevant keywords. On average, there were 9.25 keywords in each topic. For each of the above-mentioned topic, a non-negative score to a Facebook post was assigned considering the following factors: (i) length of the post, (ii) number of keywords relating to that topic, and (iii) number of keywords relating to that topic present in the post. Finally, the maximum scored topic is assigned to the post. If a post didn’t contain any keyword from any of the topics, the post was labelled as non-policy issue. Among the 6,122 posts, 2,584 were assigned to one of the policy topics.

**Sentiment Analysis**

Sentiment Analysis is a process of understanding whether a text possesses a positive or a negative sentiment, or is neutral in tone. This paper analyzed sentiments of status messages, the comments, and replies to comments. The purpose was to understand the sentiment of Facebook followers towards posts by the candidates. Sentiment of comments reflects views of the public.
towards a candidate’s opinions. There are several open-source sentiment analysis tools such as TextBlob and vaderSentiment. TextBlob offers two sentiment analysis implementations—PatternAnalyzer (based on the pattern library) and NaiveBayesAnalyzer (an NLTK classifier trained on a movie reviews corpus). So, it was expected that TextBlob might work better with review type text. But the dataset used in this study required a dedicated tool that could analyze social media content. That's why "VADER Sentiment Analysis” (Hutto, 2016) was chosen. It is a Python-based lexicon and rule-based sentiment analysis tool that performs well to detect sentiments expressed on social media. Moreover, it can handle emoticons, uses of which are very common in social media. TextBlob can't handle emoticons. The VADER analysis assigns a value between -1.0 and 1.0 as the sentiment score where higher value means positive sentiment and lower value indicates negative sentiment.

Results

RQ1. Research question 1 examined the extent to which Trump and Clinton interacted directly with followers on Facebook. Of 2.8 million comments and 0.6 million replies analyzed, only 28 comments came from Trump and 43 from Clinton. Neither of the candidates participated in discussion with followers beneath their comments. The study did not find any reply from any candidate to any user comment. In other words, results show no direct interaction between a candidate and her or his followers.

RQ2. Research question 2 examined the extent to which the candidates discussed policy issues on Facebook. The automated topic classifier found that 51% of 2,412 messages contain at least one keyword relating to 14 popular policy issues. Topic scores range from 1.06 to 0.01. A total of 53.8% of Trump’s 1,233 status messages and 48% of Clinton’s 1,179 status messages contained a topic word. Of Trump’s 663 messages that mentioned policy issues, 51.7% focused
on terrorism/national security, 8.9% on foreign policy, 8.6% on equality, 6.6% on immigration, and 5.9% on economy. The least-focused issues on Trump’s Facebook page are social security, abortion and environment. Of Clinton’s 566 messages that mentioned policy issues, almost 35% focused on terrorism/national security, 14.7% on equality, 9.9% on LGBT, 9% on economy, and 4.4% on immigration. The least-focused issues on Clinton’s page are veteran, environment and social security. The results indicate that both candidates focused heavily on terrorism and national security, but Trump talked about this topic more often than Clinton. Social security and environment appeared to be the least important topic on both candidates’ pages.

Table 2: Number of Posts and Reaction Statistics by Topic

<table>
<thead>
<tr>
<th>Topic</th>
<th>Average Number of Posts</th>
<th>Average Comments per Post</th>
<th>Average Shares per Post</th>
<th>Average Reactions per Post</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trump</td>
<td>Clinton</td>
<td>Trump</td>
<td>Clinton</td>
</tr>
<tr>
<td>Abortion</td>
<td>1</td>
<td>34</td>
<td>12811</td>
<td>2023</td>
</tr>
<tr>
<td>Economy</td>
<td>77</td>
<td>134</td>
<td>4974</td>
<td>3372</td>
</tr>
<tr>
<td>Education</td>
<td>19</td>
<td>72</td>
<td>4351</td>
<td>3296</td>
</tr>
<tr>
<td>Environment</td>
<td>5</td>
<td>22</td>
<td>4017</td>
<td>3181</td>
</tr>
<tr>
<td>Equality</td>
<td>119</td>
<td>207</td>
<td>9201</td>
<td>3902</td>
</tr>
<tr>
<td>Foreign policy</td>
<td>84</td>
<td>55</td>
<td>6679</td>
<td>3164</td>
</tr>
<tr>
<td>Gun policy</td>
<td>15</td>
<td>18</td>
<td>5341</td>
<td>1530</td>
</tr>
<tr>
<td>Healthcare</td>
<td>26</td>
<td>49</td>
<td>5613</td>
<td>1837</td>
</tr>
<tr>
<td>Immigration</td>
<td>71</td>
<td>66</td>
<td>5845</td>
<td>2058</td>
</tr>
<tr>
<td>Social security</td>
<td>6</td>
<td>12</td>
<td>7342</td>
<td>2373</td>
</tr>
<tr>
<td>Supreme court appointments</td>
<td>11</td>
<td>25</td>
<td>4723</td>
<td>2514</td>
</tr>
<tr>
<td>Terrorism/National Security</td>
<td>673</td>
<td>450</td>
<td>9445</td>
<td>3523</td>
</tr>
<tr>
<td>Trade policy</td>
<td>65</td>
<td>31</td>
<td>3751</td>
<td>8186</td>
</tr>
<tr>
<td>Treatment of LGBT</td>
<td>7</td>
<td>164</td>
<td>5084</td>
<td>2003</td>
</tr>
<tr>
<td>Veteran</td>
<td>43</td>
<td>23</td>
<td>6526</td>
<td>3768</td>
</tr>
<tr>
<td>Other</td>
<td>1812</td>
<td>1726</td>
<td>10245</td>
<td>3548</td>
</tr>
</tbody>
</table>

The table shows engagement (comments, shares, reactions) statistics of the Facebook posts relating to each of 14 topics for each candidate.
Though over half of the status messages mentioned keywords relating to policy issues, the number of messages that clearly indicate position of a candidate on a policy is a lot smaller. A manual coding of the messages by the researchers found that only 10.2% messages of Donald Trump clearly indicate his position on a policy, compared to 17.1% of Clinton’s messages.

RQ3. Research question 3 examined the extent to which the candidates used Facebook for building image. A majority (60.5%) of the total 6,122 posts contained original images--photos and videos--featuring the candidates (see Table 3). Trump attached images to 59.3% of posts and Clinton to 61.5% of posts. Analyses of two other types of status messages--message only and message with links--show that both candidates published more messages using image building techniques than discussing policy issues. Trump used image building techniques in 89.8% messages and Clinton used 82.9% messages for this purpose. Of total 1,233 messages of these two types published by Trump, 46.1% focused on strategies highlighting his personal and his campaign’s successes, 31.6% focused on attacking and criticizing Clinton and her campaigns, and 10.8% focused on personalizing Trump.

Table 3: Engagement on Status message by image-building technique

<table>
<thead>
<tr>
<th>Technique</th>
<th>Number of Posts</th>
<th>Average Comments per Post</th>
<th>Average Shares per Post</th>
<th>Average Reactions per Post</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trump</td>
<td>Clinton</td>
<td>Trump</td>
<td>Clinton</td>
</tr>
<tr>
<td>Strategy</td>
<td>569</td>
<td>532</td>
<td>6464</td>
<td>1795</td>
</tr>
<tr>
<td>Personalization</td>
<td>133</td>
<td>105</td>
<td>5495</td>
<td>2085</td>
</tr>
<tr>
<td>Negative Campaign</td>
<td>390</td>
<td>222</td>
<td>7016</td>
<td>3670</td>
</tr>
<tr>
<td>Policy Issues</td>
<td>126</td>
<td>202</td>
<td>5526</td>
<td>1400</td>
</tr>
<tr>
<td>Other</td>
<td>14</td>
<td>117</td>
<td>4461</td>
<td>1454</td>
</tr>
</tbody>
</table>

The table shows engagement (comments, shares, reactions) statistics of the Facebook posts relating to each image-building technique used by each candidate.
Clinton used 45.1% of her messages for strategic purpose, 18.9% for attacking Trump, and 8.9% for personalization. The results showing both candidates used their Facebook pages mostly for building image, support the concept of *discursive normalization*.

**RQ4.** Research question 4 asked “To what extent, do followers approve or disapprove Facebook status messages of the 2016 U.S. presidential candidates?” The sentiment analysis model found that Trump received more positive comments (about 0.8 million) than Clinton (about 0.65 million). And Clinton received more negative comments (about 0.65 million) than Trump (about 0.42 million). The researchers also performed a sentiment analysis focusing the topic of the posts. For example, Clinton posted 57.46% of 134 positive posts while Trump posted 59.74% positive posts. There is no significant difference in other sentiment level for this topic. One important observation is that, although Trump posted more negative posts than the positive ones, he got more positive comments.

**Figure 1:** This figure shows the number of posts at different sentiment level. For example, more than 1800 posts published by Donald Trump are judged as positive, whether the number is 1500 for Hillary Clinton.
To find out more proof of selective exposure, the researchers looked for unique users who participated in discussion by posting comments to posts. There was a total of 848,767 unique commenters on the two pages combined. Trump got a total of 523,132 unique commenters while Clinton got 379,305 unique commenters. There were 53,670 unique users (6.32% of total unique users) who commented on both pages. This indicates that a great majority of the Facebook followers of Donald Trump and Hillary Clinton (96.68%) did not participate in any discussion on the page of the candidate they didn't like. This shows that selective exposure exists on social media when it comes to politics.

RQ5. Research question 5 asked “Which topics discussed in Facebook status messages of the 2016 U.S. presidential candidates attracted more engagement?” To address this question, the authors analyzed topic distribution of the posts assessed by a dictionary-based topic classifier along with the engagement records the posts generated. Three matrices were used to measure engagement—comments, shares and reactions (see Table 2 on page 18). Note that on Facebook, reaction is defined as the expression of Like, Haha, Wow, Sad, Angry and Love.

Table 2 (page 18) shows the average number of comments, shares and reactions per post for each topic. In general, Donald Trump’s posts had more engagement than Hillary Clinton’s. Two exceptions in terms of the number of comments are: trade policy and veteran. Another interesting observation is the posts that don’t belong to any popular topic (the Other category) also generated a lot of engagement. It shows, there is no clear evidence that the presence of popular topics in the Facebook posts attract more engagement.
Figure 2: This figure shows the number of reactions generated per 10,000 followers was higher on Trump’s page than Clinton’s page.

To understand the relative activeness of the followers in each campaign page, the number of comments, shares, reactions produced per 10,000 followers over time was also analyzed. The reason, the authors looked at numbers per 10,000 is because the two pages have very unequal number of total followers. Figures 2 and 3 (next page) show the number of reactions and shares generated per 10,000 followers. In general, there is a higher number of comments and reactions generated per 10,000 followers in case of Donald J. Trump than of Hillary Clinton.
Figure 3: This figure shows a comparison of shares on two pages.

Discussion

Much of the research on election campaigns on social media focuses on new opportunities (e.g., connectivity and interactivity) that social media provides to engage voters in discussion on civic issues. This interdisciplinary study addressed some challenges to civic engagement on social media, Facebook in particular. It dived deep into the Facebook pages of Donald Trump and Hillary Clinton to test the theory of *Normalization of the Cyberspace* that suggests that online campaign techniques and voter engagement are similar to that of real life. This study provides a deep understanding of how candidates of the 2016 U.S. presidential election interacted with followers on Facebook, and demonstrates the extent to which engagement patterns on this social networking site resemble offline engagement patterns. The
study used a machine learning model for sentiment analysis and an automated topic classifier to analyze 3.4 million comments and replies on 6,122 status messages and found evidence of normalization on the Facebook pages of Trump and Clinton.

This study largely supports existing works on online political engagement, which suggest that online interaction between political elites and ordinary citizens follows the traditional top-down model (Lin, 2015; Nielsen & Vaccari, 2013; Schweitzer, 2011; Sweetser & Lariscy, 2008). The results show that neither of the candidates directly interacted with any follower by responding to any comment. Both candidates posted a small number of comments beneath their status messages, but those comments were meant to remind followers of something, not to start a discussion. Many status messages on both pages generated long threads of discussion, but the interactions were all among the followers themselves. As Nielsen and Vaccari suggested, even small-scale direction communication between voters and candidates is non-existent on these pages.

The study also supports the literature that suggests candidates use social media to disseminate information mainly to build and maintain image online, not necessarily to engage voters in discussion on policy issues (Bimber, 2014; Davis et al., 2010; Hurcombe, 2016). Data shows that over 60% of posts on Trump and Clinton pages combined contained original photos or videos—tools for building images. A vast majority of the remaining status messages was also aimed at building images of a candidate or her/his campaign and attacking the opponent. Trump used nearly 90% of posts (excluding photos and videos) to disseminate information strategic to his campaign or to attack the opponent. Clinton used over 83% of her status messages to do the same. Compared to the amount of status messages used for image-building techniques, policy issues got very little attention. Of the status messages brought up in the messages, a small
number provided details. As the automated topic detection classified identified, nearly half of the status messages contained at least one keyword relating to a policy issue. Manual coding by authors shows a vast majority of those messages lacked depth.

It may be cautiously said that the data supports the idea that only supporters and fans of a candidate follows her/his social media pages and engage in discussion with like-minded individuals (Davis et al., 2010; Norris, 2003). The sentiment analysis indicates that sentiment of followers’ comments on both pages are mostly positive. However, some comments contained negative sentiment, which may indicate existence of interactions among people from opposite camps. The data shows that a little over 6% unique users commented on both pages, meaning over 93% commenters participated in discussion on only one page. The reaction data provides a better understanding of this argument. Many followers on both pages reacted to candidates’ posts with emoticons (e.g., sad, angry, wow, haha). In other words, the data suggests that some interactions take place among followers from different camps (e.g., Democrats, Republicans, and Neutral) although interactions are primarily among like-minded individuals.

The findings of the study are largely consistent with the theory of the normalization of the cyberspace regarding both relational normalization and discursive normalization (Schweitzer, 2011). The data shows that typical communications practices—top-down communication model, uses of social media to build and maintain image, and selective exposure—are evidently present on the pages of both Trump and Clinton. The findings of this study imply that Facebook works mainly as an internal tool for a candidate and her/his support base. The candidates use it to build image and disseminate information to supporters. Facebook is not a place where candidates and voters interact directly. Nor may it work as a place where candidates are able to convert voters. However, there are some rays of hopes for civic
engagement. Facebook facilitates some interaction among voters from opposing camps. However, the extent to which this interaction helps improve voters’ awareness of various aspects of policy issues remains the topic for future studies.

The manual coding of Facebook posts in terms of used campaign technique adds significant value to the future development of automated campaign technique detection. Specifically, the coded data can act as a ground-truth for supervised learning models. This will facilitate analyzing campaigns of other English-speaking candidates. This potentially can also benefit other computational journalism (Cohen et al., 2011) works, such as automated fact-checking, targeted advertising and developing next-generation campaign strategies.

**Contributions, Limitations and Suggestions for Future Studies**

This timely study adds to the literature of political communication, election campaigns in particular, by testing some important aspects of an important theory. To do so, it analyzed a large data set with multiple methods. A study of this scale examining election campaign on Facebook is rare in this field. It combined the latest methods of analyzing big data with traditional content analysis method to get a deep understanding of the topic. The authors also developed a dictionary-based topic detection model to classify policy issues in status messages, which would be a useful tool for researchers in the future. It analyzed statuses, comments and reactions in several ways, which added depth to the study.

The study has several limitations as well. First, replies to comments, which account for a large part of interaction on Facebook, were not included in the analysis. Second, the study used a key-word based model developed by the authors for topic detection. Adding an external dictionary could have performed better. Third, the study used an existing model of sentiment analysis. Developing a stance detection model and training it on comments on election campaign
pages would have yielded better results. Fourth, our coding is solely based on the post text and the comment. Sometimes, a post can come with a link or video or photo which bring important context. As these contexts were not considered, our coding has some room for improvement.

Several of these limitations might be considered opportunities for future research. Researchers can develop stance detection model only to study comments on social media pages of election campaigns. Researchers can also include sub-comments in analysis, which might provide better insights into civic engagement. Researchers can also use the methods used in this study to further test the theory of the normalization on other forms of online political communication. Future study can also focus on refining the theory.
References


Da Silva, M. T. (2013). Online forums, audience participation and modes of political discussion: readers' comments on the Brazilian presidential election as a case study/Foros en línea, participación del público y modos de la discusión política: comentarios de los lectores
sobre las elecciones presidenciales de Brasil, un caso de estudio. *Comunicación y Sociedad*, 26(4), 175.


