Journal of Universal Computer Science, vol. 23, no. 3 (2017), 236-255
submitted: 15/9/16, accepted: 1/3/17, appeared: 28/3/17 © J.UCS

Generating Politician Profiles based on Content Analysis of Social Network Datasets

Klara Grčić
(University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia
klara.grcic@fer.hr)

Marina Bagić Babac*
(University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia
marina.bagic@fer.hr)

Vedran Podobnik
(University of Zagreb, Faculty of Electrical Engineering and Computing, Zagreb, Croatia
vedran.podobnik@fer.hr)

Abstract: Social networks are nowadays an influential tool in the hands of the centres of political power because of their possibilities for direct and two-way communication with citizens in real time, dissemination of information, or a self-promotion and marketing. The use of social networks in the political context has become extremely important in the analysis and prediction of elections and generally in monitoring activities of politicians and public opinion. In this paper, we provide a content analysis of Facebook activities of leading European Union (EU) politicians to generate their extended individual profiles. Based on these profiles, a comparative analysis between the European Commissioners (i.e., EU ministers) and Croatian ministers is provided showing certain unexpected differences in their online behaviour. Summarizing these results, a model for prediction of online political behaviour is proposed.

Keywords: Social media; Facebook; content analysis; politicians; European Commission, Croatia
Categories: H.3.1., I.2.7, J.4, K.4.1, K.4.2, K.4.3

1 Introduction

Facebook today has more than 1.6 billion monthly active users and still a 15% yearly increase [Facebook Reports, 16]. Because of that large number, Facebook is easiest and cheapest way to reach a large number of people. This was also recognized in politics, especially since 2008 when the US presidential elections were held. Facebook had become so popular in politics that CNN posted a question: "Will the 2008 US election be won on Facebook?" [Rawlinson, 07]. Many analysts agreed that Obama's activities on Facebook helped him win the elections because he could easily reach a large number of young people [Dutta, Fraiser 08]. The results of a study focusing on the 2008 presidential elections [Woolley et al. 10], which was based on a quantitative content analysis of more than thousand Facebook group pages about

* Corresponding author: Marina Bagić Babac; University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, HR-10000 Zagreb, Croatia; marina.bagic@fer.hr.

Barack Obama and John McCain, showed that Obama was characterized more positively through these Facebook groups.

While most of the research analysed usage of social media during elections, there is also a paper that showed how politicians use Facebook and Twitter on a regular basis [Larsson, Kalsnes 14]. The results indicate that politicians of national parliaments of Norway and Sweden post usually one status or tweet per day.

When analysing comments of Facebook users on political posts and connection between gender and online political participation, the results demonstrated that men post more negative comments both to parties and to non-political Facebook users [Vochocová et al. 16]. Another study focused on supporters' messages on Facebook, exploring how the Tea Party supporters used Facebook. It has shown that two themes – *attack* and *encouragement* – are the most important to produce an online social identity [Morin, Flynn 14].

One of few studies which explored the content of politicians' Facebook posts deals with the US 2008 and 2012 presidential candidates and shows that some candidates used Facebook to express *fear* or *anger*, while others used *humour* and *enthusiasm* [Borah, 16]. As it had been expected, the statuses with humour and enthusiasm often got more likes and shares.

Based on the analysis by [Golbeck et al. 09], it is shown that Congress members are also largely using Twitter to communicate the same type of information their offices would share in other media. Twitter is more used by politicians and because of that more research was made to analyse their activities on Twitter than on Facebook. Scholars who were examining the political impact of social networking sites through quantitative analysis of Twitter usage by Australian politicians, concluded that politicians who use Twitter to communicate usually achieve more political benefit [Grant et al, 10]. They also compared Australian politicians' Twitter profiles with random Australian Twitter users and found out that politicians tweet significantly more than Australians in general. Furthermore, there are studies [Bakliwal et al. 13] discussing sentiment analysis of political tweets which support only three possibilities – *positive*, *negative* or *neutral* – how tweets could be classified, as well as analyses in which young adults had to choose descriptions for each status to find out if they recognize different types of posts talking about divergent topics [Vraga et al. 16]. In addition, researchers who made automated content analysis, came to the conclusion that average costs could be reduced for analysing large collections of text. The only handicap is that automated content analysis can never substitute careful reading [Grimmer, Stewart 13].

The other study tried to perceive how people use Twitter for political purposes [Bode, Dalrymple 14] and came to a conclusion that political Twitter users are very interested in politics and usually enough wealthy to donate campaigns. Shortly, a Twitter usage during various elections became so popular that researchers started to investigate if tweets had become predictive. However, a study reported that tweets were more reactive than predictive [Murthy, 15]. Furthermore, social network analysis could be used to learn about anthropological and sociological aspects of modern social movements, like it was done for the Iran's Green Movement [Khonsari et al. 10].

According to some scholars, politicians have three main motives for using social media – *marketing*, *mobilization* and a possibility for *communication* with voters

[Enli, Skogerbø 13]. Additionally, users get a feeling of connectedness while using status update messaging. That feeling of being connected is associated with the amount of messages and not with the type of the exchanged information [Köbler et al. 10].

In this paper, we focus on statuses posted by politicians, specifically, the European Commissioners (i.e., EU ministers) and Croatian ministers. The idea was to categorize these statuses based on their content with the goal to find out which are the topics these politicians write the most about as well as if there are differences between statuses written by European and Croatian politicians.

Collected statuses were the ones posted in the first four months since the organizations were formed. They were read carefully and assigned to one or more of the following categories: *Information*, *Past events*, *Current actions*, *Future plans* and *Personal messages*. Following, the comparison between the number of Facebook *likes*, *shares*, and *comments* on different categories from different datasets was made, as well as statistical tests to find out if the difference between these variables was statistically significant. Summarizing these results based on content categories, a model for prediction of online political behaviour is proposed and tested.

2 Research Questions

This study aims to answer the following questions relating to content analysis of Facebook posts of European Commissioners and Croatian ministers;

- *RQ1: What kind of posts appeared at European Commissioners' and Croatian ministers' Facebook pages?*
- *RQ2: How European Commissioners' and Croatian ministers' posts affect the public (in terms of Facebook likes, shares and comments)?*
- *RQ3: What are differences/similarities between European Commissioners' and Croatian ministers' posts?*

Based on the theoretical framework, and the extended individual politicians' profiles obtained by content analysis, we state the following hypothesis with its sub-hypotheses:

Hypothesis H: *When writing about past (H1), present (H2), or future (H3) on Facebook, European Commissioners and Croatian ministers affect the public similarly, i.e. their posts receive similar amount of: likes (a); shares (b); and comments (c).*

3 Data Collection and Methodology

The preparatory work for this analysis was to find out which European Commissioners and Croatian ministers have official Facebook pages. If a politician had a Facebook profile, her/his profile name was used to gather information, i.e., her/his statuses along with the number of *likes*, *comments*, and *shares* for each status. Furthermore, each status was coded to one or more predefined categories, according to [Golbeck et al. 10]: *Information*; *Current action*; *Future plans*; *Past events*; or

Personal messages. In addition, there were cases when statuses were not added to any category of these but in one or both of two additional categories: *links to articles*, and *photos/videos*.

Based on characteristics of the variables (binary), expected distributions across coding categories (binomial), and the number of coders (three), the Krippendorf's alpha [Krippendorff, 11] is chosen as an appropriate index of inter-coding reliability in addition to average percent agreement, which both indicated that the coding procedures were reliable (Table I). Three coders were trained as a group using a series of examples, and disagreements among them were resolved using a majority rule [Lombard et al, 10].

Coding variable	**Average percent agreement**	**Krippendorff's alpha**
Information	.79	.73
Current action	.84	.79
Future plans	.90	.85
Past events	.95	.90
Personal messages	.96	.91

Table 1: Reliability statistics for coding of Facebook posts

Here are few examples of coded posts;

- *Informative* post about *Future plans*:

"*We've decided at Foreign Affairs Council to extend existing sanctions and support diplomatic efforts to implement Minsk agreement #Ukraine*" by the European Commissioner Federica Mogherini.

- *Informative* post about *Current actions*:

"*Meeting Party of European Socialists friends and colleagues ahead of the European Council meeting in Brussels today. Will discuss European Commission investment plan and situation in Ukraine*" by the European Commissioner Karmenu Vella.

- *Informative* post about *Past events*:

"*Sir Winston Churchill in The Hague, 1948. His speech on that occasion is still worth reading*." by the European Commissioner Frans Timmermans.

- *Non-informative* post with a *Personal message*:

"*A happy 2015 to all. May the new year bring peace.*" by the European Commissioner Federica Mogherini.

For Croatia, statuses were collected for the period since 23 December 2015, when the Prime Minister of Croatia was named, till 30 April 2016 (when this analysis began). The other dataset, used to compare with Croatian politicians' statuses, were statuses gathered from European Commissioners. It was important to choose the same relative period (four months since forming the Commission) for the results to be more relevant. Because of that, dates relevant for European Commissioners' statuses are from 1 November 2014 till 1 April 2015.

Our final dataset is publicly available at *goo.gl/dSJbfk*.

3.1 Dataset Overview

In Figure 1, a bar chart with the percentage of Facebook profiles can be seen. The first bar shows the percentage of Croatian ministers who (do not) have official Facebook profiles, and the second one shows the percentage of European Commissioners who (do not) have official Facebook profiles. Both values are similar (around 50%), with Croatian ministers having a slightly higher percentage of Facebook profiles.

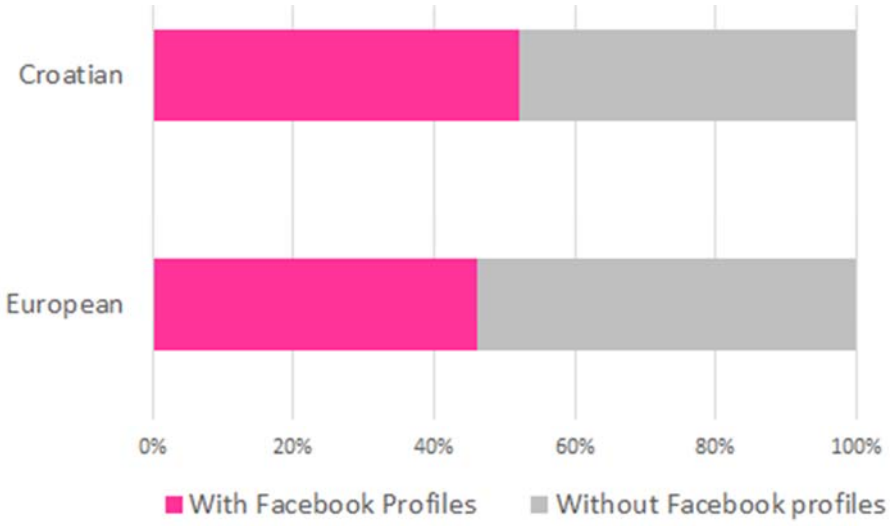


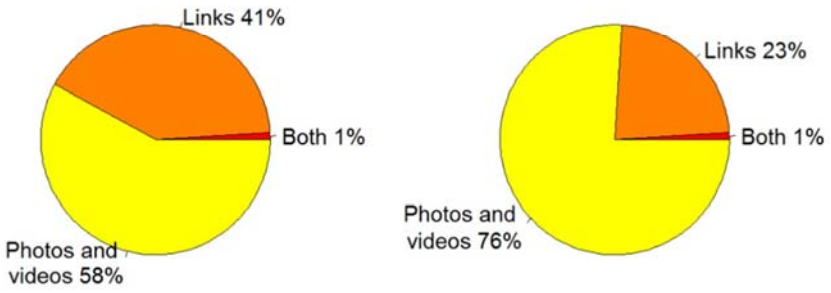
Figure 1: Percentage of Croatian ministers and European Commissioners who have Facebook profiles

Utilising Facebook's Graph API Explorer, we retrieved 800 statuses written by Croatian ministers and 1,750 statuses written by European Commissioners, where an average number of statuses per profile of Croatian minister is 50, and an average number of European Commissioners is 102. It is important to note that a significant number of Croatian ministers did not even have Facebook pages until they have become ministers.

Besides the statuses that fit into one of more categories, there were also posts which did not include any textual message, but only a *link* and/or *photo/video*. Among Croatian politicians' statuses, there are 137 posts of this kind, which makes about 17% of all their statuses. The case is similar when looking at European politicians. They have 166 statuses without any textual message out of total of 1,067 statuses written in English (non-English statuses were filtered out). That makes around 15% of their statuses written in English.

In addition, any status can be a *link*, or a *photo/video* or both. Figure 2a shows the rate of these categories among statuses without any textual message. It can be seen that European Commissioners post both links and photos/videos, with percentages of 41% for links and 58% for photos/videos. Only 1% are statuses with both a link and a photo/video. Looking at Figure 2b, it can be seen that percentage of statuses with both

a link and a photo/video is the same, however Croatian ministers post more photos and videos than European Commissioners.



a) *b)*

Figure 2: The ratio of categories of statuses without any textual message posted by: a) European Commissioners; b) Croatian ministers

One of the reasons for these results for the non-message posts is because everything that European Commissioners do is published on the official page of European Commission, and then they often post a link on Facebook. There could also be found videos of their speeches which they post as links. In Croatia, situation is different – while there is as well the official web page which posts news about the Croatian government, Croatian ministers generally do not publish these news on their Facebook profiles.

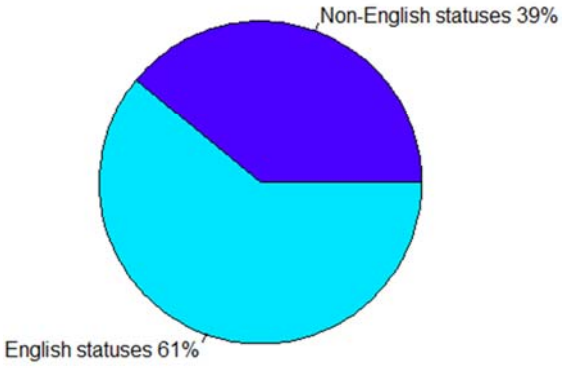


Figure 3: Percentages of statuses written in English and completely non-English statuses

All statuses written by Croatian ministers are in Croatian. When it comes to European Commissioners' statuses, the situation is not that simple. Most of the statuses were written in English or bilingual, usually in English and in a language of the country from which the Commissioner comes from. There were also situations where authors had written only in their mother tongue. The ratio between English and wholly non-English statuses can be seen in Figure 3. Non-English statuses of European Commissioners were not included in the analysis presented in this paper.

3.2 The Use of Language in Facebook Statuses

Another aspect of statuses' content are words that appeared most frequently in European Commissioners' and Croatian ministers' statuses. All non-English words written by Commissioners were excluded.

Words	***Frequency***
European	381
today	164
commission	161
meeting	132
Europe	130
new	109
energy	95
union	82
minister	80
investment	74
good	71
support	67
great	66
parliament	66
work	65
day	64
people	61
council	58
president	55
regional	51
years	50
US	49

Table II: The most frequent English words in European Commissioners' Facebook posts

Table II shows the most used English words written by European Commissioners along with number of appearances for each word. It can be seen that words "Europe" and its form "European" are among most mentioned words. Furthermore, words "today" and "meeting", used to tell about their plans for that day, are frequently written by Commissioners as well. It also can be noticed that positive words such as "good", "great" and "support" have high frequency as well.

On the other hand, Table III reveals the most frequently used Croatian words (with their translations in English) written in statuses by Croatian ministers. The situation is similar to Commissioners' statuses – the word "Croatia" ("*Hrvatska*") and different forms of "Croatian" ("*hrvatske*", "*hrvatskoj*") are very common. However, comparing to Table II, it can be seen that Croatian ministers do not use such as positive words as European Commissioners but more formal language.

Words	***Meaning***	***Frequency***
hrvatske	*Croatian*	177
Hrvatska	*Croatia*	126
vlade	*government*	114
EU	*EU*	104
ministar	*minister*	104
hrvatskoj	*Croatian*	101
republike	*Republic*	93
godine	*years*	80
vlada	*government*	69
poslova	*affairs*	60
hrvatsku	*Croatia*	59
dana	*days*	48
politike	*policy*	45

Table III: The most frequent Croatian words in Croatian ministers' Facebook posts

4 Empirical Results of Content Analysis

4.1 European Commissioners' Statuses

In answer to the first (RQ1) and second (RQ2) research questions relating the content and effects of the posts, the results of content analysis are shown here. Bar plot in Figure 4 shows all categories with percentage of statuses which belong to each category. The *Information* is the most common, with around 75% share. All other categories do not have even close number of statuses. *Current actions* are about 24% of statuses, while about 16% of statuses provided certain *Personal message* from European Commissioners. Two least common categories were *Future plans* and *Past Events*, both with less than 10%.

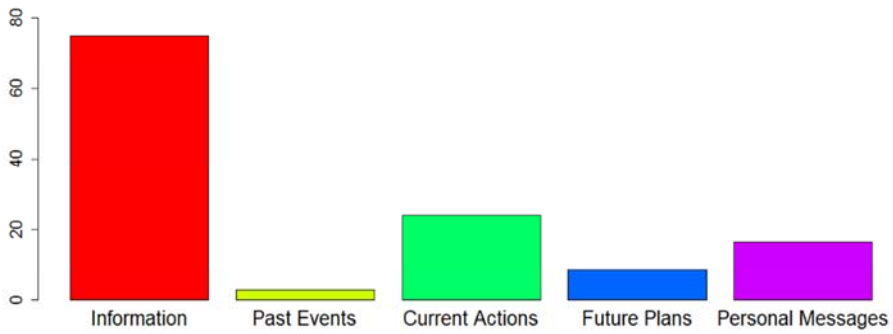


Figure 4: Frequency distribution (percentages) of European politicians' statuses by content values

In the Figure 5a, a summary statistics of number of *likes* per post for three categories – *Past events*, *Current actions*, and *Future plans* – is shown. The most liked statuses are those from the *Past events* group, while statuses belonging to group *Future plans* get slightly more *likes* than the one from *Current Actions*.

Looking at the Figure 5a, it can be noticed that the biggest outliers for *Current actions* and *Future plans* have equal number of *likes* per post. Actually, it is the same status, posted by Kristalina Georgieva, European Commissioner for Budget and Human Resources. She is Bulgarian and most of her statuses were written in both Bulgarian and English, just like this one. Given that this post is about Bulgaria and the investment Bulgaria got, the majority of people who liked it are Bulgarians;

"*Days after with Jyrki Katainen we successfully negotiated the #EFSI, #Bulgaria joins with 100 mln euro. I am proud to see my country among the first! I expect that this contribution will help Bulgaria attract private capital, achieving a significant multiplication effect on its investment and implement projects which will change the lives of Bulgarian citizens for the better. #InvestEU*"

The two biggest outliers from the *Past events* category are also statuses posted by Kristalina Georgieva. The one with more (i.e., 3,539) likes is:

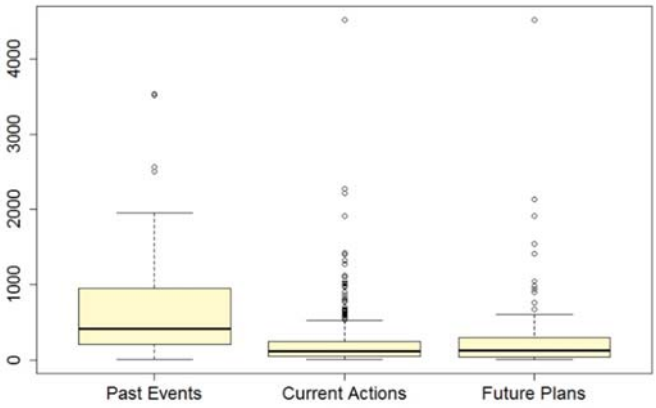
"*First time at Dimitar Peshev Plaza in Washington DC with the Bulgarian Ambassador to the USA, Ms Elena Poptodorova. Dimitar Peshev saved thousands of Bulgarian Jews from deportation. The Plaza in Washington DC is named after him.*"

This is also a status liked mainly by Bulgarians because it was about Dimitar Peshev, who rebelled against pro-Nazi cabinet and saved thousands of Bulgaria's Jews by preventing their deportation.

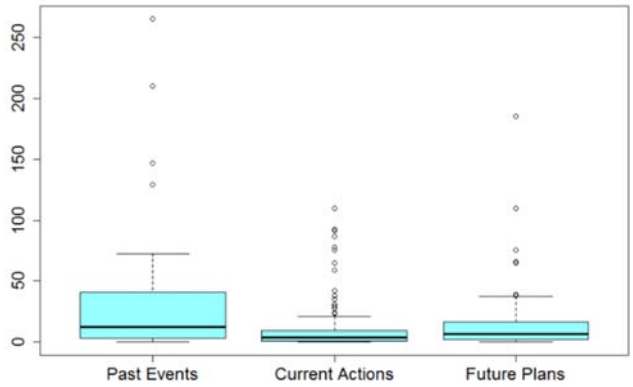
The second one has the completely different topic:

"*Remembering my first post on Facebook 5 years ago. Pictures from my first hearing at the European Parliament. http://on.fb.me/1Jtws3B*"

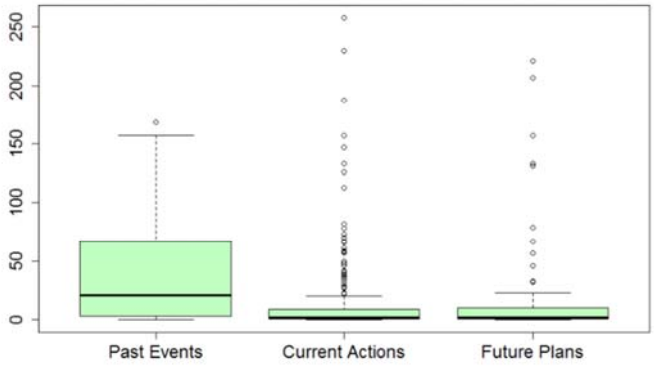
It is the status liked by 3,516 people, and in this status Kristalina Georgieva was remembering on her first post on Facebook, published 5 years before.



a) Numbers of likes



b) Numbers of shares



c) Number of comments

Figure 5: Summary statistics of European politicians' statuses by content categories: numbers of a) likes; b) shares; and c) comments

Figure 5b, describing European Commissioners' status *shares*, is quite similar to Figure 5a. One of the differences is in the outliers. Two greatest outliers are statuses belonging to the *Past events* group. The biggest outlier is again for the status written by Kristalina Georgieva:

"*25 years ago today began Bulgarian transition to democracy. One generation later it is time for the transition to end.*"

It was shared 265 times and in this status Kristalina Georgieva remembered the Bulgaria's democratic revolution which lead to end of the People's Republic. Also, the first multi-party elections were held and country's name was to change to the Republic of Bulgaria. Therefore, it is not surprising that this was one of the most shared statuses, considering this was important event in the Bulgarian history.

Figure 5c describes European Commissioners' status *comments*. Although there are many statuses with only few comments or none, there are statuses with more than 250 comments as well. This is the reason the boxplots for *Current action* and *Future plans* are characterised with small medians and a lot of outliers. The most commented status is the one written by Jean-Claude Juncker, the European Commission President, assigned to the *Current actions* category:

"*In today's College meeting we spoke about transparency - EU citizens have the right to know with whom Commissioners and Commission staff meet in the context of the legislative process*."

The only outlier from the *Past events* category refers to status written by Frans Timmermans, the First Vice President and Commissioner for Better Regulation, Inter-Institutional Relations, Rule of Law and Charter of Fundamental Rights:

"*Heerlijk weer voor een wandeling door Aken. Strolling through Aachen with Max and my mom, where she reminisces about her trips to Aachen in her youth not long after the war. Passport, thorough security check at the border to catch people smuggling coffee into Germany and the hassle of changing guilders into D-Mark. Sound like ancient history to Max*."

This is also a personal status, because Frans Timmermans was talking about his family and his mother's travel memories which seemed strange to his son Max.

The most shared status from *Future plans* category was again posted by Jean-Claude Juncker. It was one of his first statuses, written on the day the European Commission was formed:

"*Europe's challenges cannot wait. As of today, my team and I will work hard to deliver the new start we have promised. http://europa.eu/rapid/press-release_IP-14-1237_en.htm*"

4.2 Croatian Ministers' Statuses

Figure 6 shows that Croatian ministers in their statuses often provided information – more than 60% of statuses were assigned to this group. All other categories have a much smaller percentage. About 20% of statuses were in the *Current action* and the *Personal message* group and a little less in *Future plans*. The topic about Croatian ministers wrote the least was *Past events*.

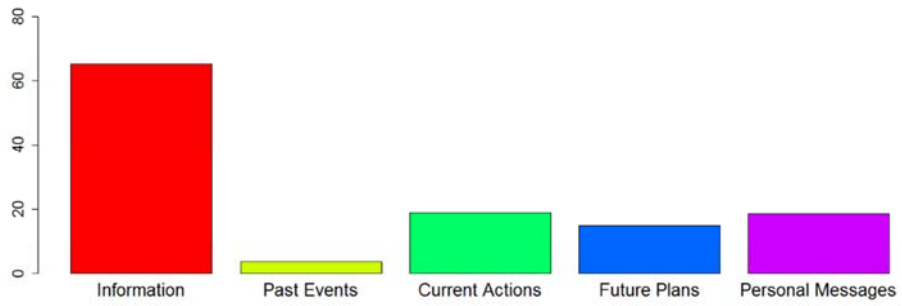
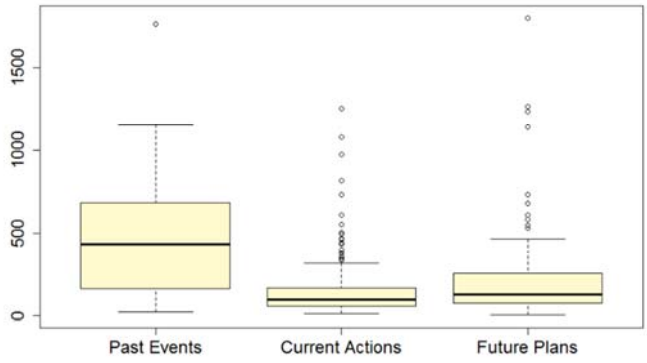


Figure 6: Frequency distribution (percentages) of Croatian ministers' statuses by content values

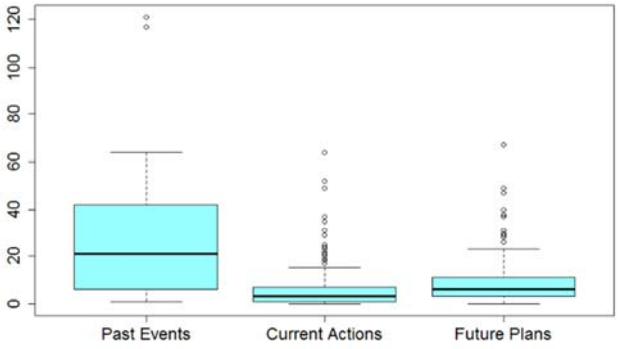
Looking at Figure 7a, it is evident that people like most those statuses written about some events happened in past. On the other hand, statuses related to some current actions are least liked and have smallest outliers as well. Looking at the Figure 7b, it can be seen that situation is similar for the number of shared statuses, but numbers are much smaller here. In addition, Figure 7c shows number of comments. Comparing the number of comments to the number of likes and shares, a situation is a bit different. The number of likes and shares always means something positive as people share someone's statuses only if they agree with them. But when it comes to commenting, someone's comment could be a sign of both agreement or disagreement. Because of that, this graph, unlike the previous graphs, is a bit ambiguous. However, considering the previous graphs, it could be easily seen that graphs are similar. As statuses are gathered from politicians' fan pages, the most people who read and comment those statuses are ones who liked the page as well.

4.3 European Commissioners vs. Croatian Ministers

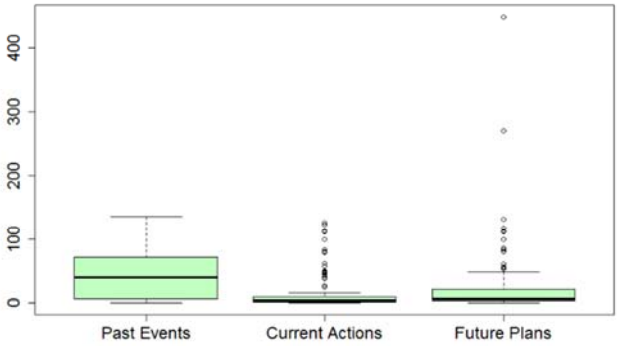
In answer to the third research question (RQ3), relating posts comparison, we have used the two sample t-test to find out if two sets of data are much different from each other. The list of test outcomes, i.e. the 95% confidence interval ($\mathbf{CI}_{.95}$) of the difference between two population means is provided in Tables IV-VI along with *t* and *p* values;



a) Numbers of likes



b) Numbers of shares



c) Numbers of comments

Figure 7: Summary statistics of Croatian ministers' statuses by content categories: numbers of a) likes; b) shares; and c) comments

Content category	$CI_{.95}$	t	p
Past event	(-809.11, -26.69)	-1.89	0.0657
Current action	(-161.98, -36.33)	3.10	0.0021
Future plans	(-233. 45, -35.02)	-1.46	0.1459

Table IV: Test results for the number of likes on statuses

Content category	$CI_{.95}$	t	p
Past event	(-39.11, -39.65)	-0.71	0.4813
Current action	(-3.51, -8.27)	-0.79	0.4276
Future plans	(-10.51, -0.85)	-1.69	0.0946

Table V: Test results for the number of shares on statuses

Content category	$CI_{.95}$	t	p
Past event	(-26.38, -22.51)	-0.16	0.8746
Current action	(-5.64, -6.16)	0.09	0.9322
Future plans	(-5.59, -20.01)	1.11	0.2684

Table VI: Test results for the number of comments on statuses

In Tables IV-VI, only one population difference – related to the number of *likes* on statues assigned to *Current actions* – was statistically significant ($p<0.01$). This is the only sub-hypothesis (H2a) that was rejected, while the others (H1a-c, H2b-c, H3a-c) are confirmed. The meaning of the rejected sub-hypothesis is that people "liked" much more European Commissioners talking about the present than Croatian ministers talking about the present. This finding implies that people trust European Commissioners much more than Croats trust their ministers. In addition, Croats are much more into the past than the present due to their short age in democracy (since 1991).

5 Predicting Political Context from User-generated Data

In this section, we utilize our politicians' dataset to predict political context, i.e. given a political post, how likely it is to be *informative*, *personal*, or talking about the *present*, *future*, or *past*. According to results from previous section, there are no significant differences between the European Commission's and Croatian datasets. Consequently, we join them to conduct a logistic regression. The only exception to results presented in Table VII is that for predicting the *present* (*Current actions*), we use only European Commission's dataset as it is significantly different from the Croatian dataset (H2a).

In the regression model, the numerical predictor variables are numbers of *likes*, *comments* and *shares* a post receives (*Likes, Comments*, and *Shares*, respectively), and the categorical variables are those showing media resources of a post, i.e., if it

contains *photos and videos*, and/or *links to articles* (*Photos&Videos* and *LinksToArticles*) and if it is declared by type as *photo*, *status*, *video* or *link* (*TypePhoto*, *TypeStatus*, and *TypeVideo*; *TypeLink* is not visible in Table VII as it is a reference value). Therefore, the regression model is as follows;

$$logit(p) = \beta_0 + \beta_1*Likes + \beta_2*Shares + \beta_3*Comments + \beta_4*Photos\&Videos + \beta_5*LinksToArticles + \beta_6*TypePhoto + \beta_7*TypeStatus + \beta_8*TypeVideo$$

The *p* variable in the model refers to a response variable, which is here *Information*, *Past*, *Present*, *Future*, and *Personal*, and betas are variables' weights. Table VII shows the logistic regression outcomes, where beta is reported in each cell with odds ratio in parenthesis.

Predictor	Information	Past	Present	Future	Personal
Likes	-3.14e-04* (.9997)	4.27e-05 (.9999)	2.15e-04* (.9998)	-4.49e-04° (.9995)	4.56e-04*** (1.0005)
Shares	.0016* (1.0016)	.0018 (1.0018)	-.0071 (.9929)	2.52e-04 (1.0003)	.0042° (1.0042)
Comments	.0004 (1.0010)	.0014 (1.0014)	-.005 (.9949)	.0004 (1.0036)	.0019° (1.0019)
Photos & videos	.1149 (1.1218)	.5652° (1.7597)	.7956 (1.1509)	1.424 (4.1533)	-.5566 (.5732)
Links to articles	1.223* (3.3984)	.5539 (1.7400)	.7537 (.8827)	1.273*** (3.5705)	-.4585 (.6322)
Type: photo	1.271 (3.5657)	1.405 (4.1354)	.9735 (1.1996)	-.292 (.5890)	1.2949* (3.6505)
Type: status	2.143** (8.5254)	2.35** (8.6105)	.9732 (2.0954)	.6169 (1.8533)	1.0974** (2.9962)
Type: video	.3976 (1.4883)	.7311 (1.7392)	.9746 (.6374)	-1.207 (.2991)	1.0857 (2.9616)
McFadden R^2	.0313	.0477	.0286	.0273	.0888

Significance codes: 0 *** .001 ** .01 * .05 °

Table VII: Logistic regression predicting whether a post is informative, personal, or is talking about past, present, or future

According to the results presented in Table VII, numbers of *likes* and *shares* are significant indicators if a post is *informative*. In addition, a post is 3.40 times more likely to be *informative* if it contains *links to articles*, and also 8.53 times more likely to be *informative* if it is a *status*. While a number of *shares* indicates a slightly positive impact on posts being *informative*, results for the number of *likes* indicate a slightly negative impact. There does not appear to be such relationships for the number of *comments*.

It is interesting to observe that numbers of *likes*, *shares* and *comments* are all indicators of a post being *personal* meaning that people enjoy this kind of posts the

most. In addition, a post is 3.65 times more likely to be *personal* if it contains a *photo*, and also 3.00 times more likely to be *personal* if it is a *status*.

Although most of the variables are not significant indicators of the *present*, *past* or *future* event, the *type* (*status*) and *photos and videos* are indicators of a *past* event, i.e. a post is 1.76 times more likely to be about *past* if it contains both *photos and videos*, as well as 8.61 times more likely to be about *past* if it is a *status*. A number of likes indicates a slightly negative impact on posts being about *future*, and a slightly positive impact on posts being about *present*. In addition, a post is 3.57 times more likely to be about *future* if it has *links to articles*.

To assess the quality of our model, we have used 10-fold cross validation, a machine learning technique to evaluate predictive models by randomly partitioning the dataset into training and testing sets [Murphy, 12]. Our dataset is randomly partitioned into ten equal size subsamples of which a single sub-sample is retained as the validation data for testing the model, and the remaining nine subsamples are used as training data. The cross-validation process is then repeated ten times, with each of the ten sub-samples used exactly once as the validation data. We have obtained accuracy 0.70 for the *Information*, 0.83 for the *Personal messages*, 0.96 for the *Past events*, 0.89 for the *Future plans*, and 0.78 for the *Current actions*, where only European Commissioners' posts were in the training and testing sets, which altogether are encouraging results. In spite of that, McFadden R^2 parameter suggests that much more research is needed (e.g., extending the set of predictor variables) before we can draw precise and more general conclusions regarding the consequences of this development.

6 Conclusions

Among digital communication tools of the twenty-first century, social media websites provide innovative ways for politicians to connect with citizens [Lukamto, Carson 16]. The analysis of media coverage from the 2010 UK General Election demonstrates that social media are now being equated with public opinion by political journalists [Anstead, O'Loughlin 15], therefore politicians' adoption of social media should cover a variety of use intensities and purposes [Hoffman et al. 16].

This study contributes to exploring applications and case studies that analyse the usage of online social networks confirming that Facebook seems to increase the personalization of politics in news reporting [Ekman, Widholm 15]. More specifically, it provides a content analysis of extended Facebook profiles of European Commissioners and Croatian ministers. These profiles are built upon their online behaviour which is observed through the analysis whether their posts are *informative* or *personal*, or are about *past*, *present*, or *future*.

The results of this study show that politicians mostly write *informative* content on their official personal pages, which can be interpreted as news reporting. Although the *past* events is something politicians least wrote about, with ratio less than 5%, it is most *liked*, *shared* and *commented* category of their statuses. While the numbers of *shares* and *comments* are similar for all categories of both datasets (Croatian and European), the numbers of *likes* are much bigger on statuses written by European Commissioners. These assumptions were statistically tested, and even though two variables (*likes* on the *Past events* and *shares* of the *Future plans*) had a difference

between datasets, that difference was not quite statistically significant. The only variable with very statistically significant difference were *likes* on statuses from the *Current actions* category.

Utilizing these findings and proposed set of predictor variables (based on questions whether a post is *informative*, *personal*, or is talking about *past*, *present*, or *future*), we have created a prediction model for online political behaviour in terms of content categories, which has shown encouraging results for the case of European and Croatian datasets. While the model potentially leads to an automated content recognition, it is built upon these limited datasets which might not be a representative of a typical political behaviour in general, so much more country- or region-specific online political context should be explored. Here, the challenge is to bridge the gap of language barriers, and, what is even more difficult, historical discourses within various countries. Therefore, further research might extend our analysis to a larger scale, e.g., prolonging periods of data collection as well as extending data collection to additional countries [Dang-Xuan et al. 13].

Although the findings of this paper are limited to a sample of politicians' profiles, recent analyses indicate that personalized and interactive uses of online media can be beneficial to politicians [Hoffman et al. 16], potentially leading to more favourable perceptions. Therefore, even our limited findings might provide new perspectives to political and social marketing [Podobnik et al. 13].

Another limitation of our study is the fact that European Commissioners are of different political parties and we have compared them to Croatian government which consists of significantly more homogenous ministers (still, they do not belong to the same party due to fact that observed Croatian government was a coalition government). Furthermore, a possible situation is that some political parties which do not have their ministers in the government may have more Facebook activities. In addition, there might be correlations between political parties and their usage of Facebook. Comparing different governments (where different political parties could be in a single government, or which can vary from country to country) might have an impact on the presented research.

Given the increased relevance of political communication in social media, it is also important for politicians to use social media more proactively to enter into dialogs and discussion with citizens. Further studies are needed that deal more specifically with questions of interactivity with users, as the emergence of user-generated content can harvest new forms of contra-flaws [Ekman, Widholm 15]. For instance, new Facebook reactions, so called *emojis* [Guynn, 16] might expand the range of user-generated data about political content, or extend the set of predictor variables in our prediction model (e.g. a number of the *love*, *haha*, *sad*, *angry*, and *wow* emojis). In addition, textual corpus of political posts should be correlated to sentiment analysis of political text [Grimmer, Stewart, 13] and user reactions, so new perspectives on user perception of political content can be observed. The same is to be done with other media content of political posts, i.e. *photos*, or *videos* as our study indicated influence of various media types to the number of user *likes*, *shares* and *comments*.

There is much more future work that could expand the presented research. Broader target group, including governments of non-European countries would allow for more interesting results. Furthermore, the Twitter activities [Bakliwal et al. 13],

microblogging [Tumasjan et al. 10], or other online activities of the politicians could be analysed and compared in the same way. Moreover, politicians' activities on social media could be related to their personal information such as gender [Bagić Babac, Podobnik 16], age [Chan, Guo 13], level of education and profession [Hoffmann et al. 16], or sentiments [Conover et al. 11]. Additionally, the countries' GDP could be correlated to its politicians' activities on social networks or topics they write about.

Acknowledgements

The authors acknowledge the support of the research project "Managing Trust and Coordinating Interactions in Smart Networks of People, Machines and Organizations", funded by the Croatian Science Foundation under project UIP-11-2013-8813, as well as the research project "SmartSocial: Beyond Social Networking of People", funded by the Ericsson Nikola Tesla, Croatia. The Croatian Centre of Research Excellence for "Data Science and Advanced Cooperative Systems" supported the authors as well.

References

[Anstead, O'Loughlin 15] Anstead, N. and O'Loughlin, B. (2015), "Social Media Analysis and Public Opinion: The 2010 UK General Election", Journal of Computer-Mediated Communication 20(2): 204-220.

[Bagić Babac, Podobnik 16] Bagić Babac, M. and Podobnik, V. 2016. "A sentiment analysis of who participates, how and why, at social media sport websites: How differently men and women write about football". Online Information Review, 40(6): 814-833.

[Bakliwal et al. 13] Bakliwal, A., Foster, J., van der Puil, J., O'Brien, R., Tounsi, L. and Hughes, M. 2013. "Sentiment Analysis of Political Tweets: Towards an Accurate Classifier", In Proc. Workshop on Language in Social Media (LASM 2013), Atlanta, Georgia, ACL, 49-58.

[Bakker, de Vreese 11] Bakker, T. P., and de Vreese, C.H. 2011. "Good news for the future? Young people, Internet use, and political participation" Communication Research 38 (4): 451-470.

[Bode, Dalrymple 14] Bode, L. and Dalrymple, K.E. 2014. "Politics in 140 Characters or Less: Campaign Communication, Network Interaction, and Political Participation on Twitter", Journal of Political Marketing, doi: 10.1080/15377857.2014.959686

[Borah, 16] Borah, P. 2016. "Political Facebook Use: Campaign Strategies Used in 2008 and 2012 Presidential Elections", Journal of Information Technology & Politics, doi:10.1080/19331681.2016.1163519

[Chan, Guo 13] Chan, M. and Guo, J. 2013. "The Role of Political Efficacy on the Relationship Between Facebook Use and Participatory Behaviors: A Comparative Study of Young American and Chinese Adults", Cyberpsychology, behaviour and social networking 16(6): 460-463.

[Conover et al. 11] Conover, M. D., Ratkiewicz, J., Francisco, M., Gonalves, B., Flammini, A. and Menczer. F. 2011. "Political polarization on Twitter", in Proceedings of the 5th International AAAI Conference on Weblogs and Social Media, Washington, DC, AAAI Press, Menlo Park, CA, 89–96.

[Dang-Xuan et al. 13] Dang-Xuan, L., Stieglitz, S. Wladarsch, J. and Neuberger, C. 2013. "An investigation of influentials and the role of sentiment in political communication on Twitter during election periods", Information, Communication & Society 16(5): 795-825.

[Edgerly et al. 16] Edgerly, S., Thorson, K., Bighash, L. and Hannah, M. 2016. "Posting about politics: Media as resources for political expression on Facebook". Journal of Information Technology & Politics 13(2): 108-125.

[Ekman, Widholm 15] Ekman, M. and Widholm, A. 2015 "Politicians as Media Producers", Journalism Practice 9(1): 78-91.

[Enli, Skogerbø 13] Enli, G.S and Skogerbø, E. 2013. "Personalized campaigns in party-centred politics", Information, Communication & Society 16 (5): 757-774.

[Facebook Reports, 16] Facebook Reports First Quarter 2016 Results and Announces Proposal for New Class of Stock, available at: https://investor.fb.com/investor-news/press-release-details/2016/Facebook-Reports-First-Quarter-2016-Results-and-Announces-Proposal-for-New-Class-of-Stock/default.aspx

[Franch, 13] Franch, F. (2013), "Wisdom of the Crowds": 2010 UK Election Prediction with Social Media", Journal of Information Technology & Politics 10(1): 57-71.

[Gerodimos, Justinussen 15] Gerodimos, R. and Justinussen, J. 2015. "Obama's 2012 Facebook Campaign: Political Communication in the Age of the Like Button", Journal of Information Technology & Politics 12 (2): 113-132.

[Golbeck et al. 10] Golbeck, J., Grimes, J.M. and Rogers, A. 2010. "Twitter use by the U.S. Congress", Journal of the American Society for Information Science and Technology 61(8): 1612-1621.

[Grant et al. 10] Grant, W.J. Moon, B. and Grant, J.B. 2010. "Digital Dialogue? Australian Politicians' use of the Social Network Tool Twitter", Australian Journal of Political Science 45(4): 579-604.

[Grimmer, Stewart, 13] Grimmer, J. and Stewart. B.M. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts", Political analysis 21(3): 267-297.

[Guynn, 16] Guynn J. "Meet Facebook's new emoting emojis: Love, haha, wow, sad and angry", USA Today, 25 Feb 2016, www.usatoday.com/story/tech/news/2016/02/24/facebook-reactions-launch/80803468/ (accessed 3 August 2016)

[Hoffmann et al. 16] Hoffmann, C.P., Suphan, A. and Meckel, M. 2016. The impact of use motives on politicians' social media adoption, Journal of Information Technology & Politics 13(3): 239-256.

[Khonsari et al. 10] Khonsari, K.K., Nayeri, Z.A., Fathalian, A. and Fathalian, L. 2010. "Social Network Analysis of Iran's Green Movement Opposition Groups using Twitter?", International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2010, Odense, Denmark, August 9-11, 2010.

[Köbler et al. 10] Köbler, F., Riedl, C., Vetter, C., Leimeister, J.M. and Krcmar, H. "Social Connectedness on Facebook – An Explorative Study on Status Message Usage", in 16. Americas Conference on Information Systems (AMCIS), 2010, Lima, Peru.

[Krippendorff, 11] Krippendorff, K. 2011. Computing Krippendorff's Alpha-Reliability, http://repository.upenn.edu/asc_papers/43

[Larsson, Kalsnes, 14] Larsson, A.O. and Kalsnes, B. 2014 'Of course we are on Facebook': Use and non-use of social media among Swedish and Norwegian politicians, European Journal of Communication 29 (6): 653-667.

[Lombard et al, 10] Lombard, M., Snyder-Duch, J. and Bracken, C.C., 2010. Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research Projects, available at: http://matthewlombard.com/reliability/

[Lukamto, Carson 16] Lukamto, W. and Carson A. 2016. "POLITWEETS: social media as a platform for political engagement between Victorian politicians and citizens, Communication Research and Practice 2(2): 191-212.

[Maynard, Funk 11] Maynard, D. and Funk, A. "Automatic detection of political opinions in Tweets", The Semantic Web: ESWC 2011 Workshops, LNCS 7117, 88-99.

[Miller, 13] Miller, W.J. 2013. "We Can't All Be Obama: The Use of New Media in Modern Political Campaigns", Journal of Political Marketing 12(4): 326-347.

[Morin, Flynn 14] Morin, D.T. and Flynn, M.A. 2014 "We Are the Tea Party!: The Use of Facebook as an Online Political Forum for the Construction and Maintenance of in-Group Identification during the "GOTV" Weekend", Communication Quarterly 62(1): 115-133.

[Murphy, 12] Murphy, K.P. Machine learning: a probabilistic perspective, 2012. Cambridge, MA

[Murthy, 15] Murthy, D. 2015. "Twitter and elections: are tweets, predictive, reactive, or a form of buzz?", Information, Communication & Society 18(7): 816-831.

[Podobnik at al. 13] Podobnik, V., Ackermann, D., Grubisic, T. and Lovrek, I. 2013. "Web 2.0 as a foundation for Social Media Marketing: global perspectives and the local case of Croatia", in Azab, N.A. (Ed), Cases on Web 2.0 in Developing Countries: Studies on Implementation, Application, and Use, Hershey, IGI Global, 342-379.

[Tumasjan et al. 10] Tumasjan, A., Sprenger, T.O., Sandner, P.G. and Welpe, I.M. 2010. "Predicting elections with Twitter: What 140 characters reveal about political sentiment", Proceedings of the International Conference on Weblogs and Social Media, Association for the Advancement of Artificial Intelligence, 178-185.

[Vitak et al. 11] Vitak, J., Zube, P. Smock, A., Carr, C.T., Ellison, N. and Lampe, C. 2011. "It's Complicated: Facebook Users' Political Participation in the 2008 Election", Cyberpsychology, behaviour, and social networking 14(3): 107-114.

[Vochocová et al. 16] Vochocová, L., Štětka, V. and Mazák, J. 2016. "Good girls don't comment on politics? Gendered character of online political participation in the Czech Republic", Information, Communication & Society 19(10): 1321-1339.

[Vraga et al. 16] Vraga, E.K. Bode, L. Smithson, A.-B. and TrollerRenfree, S. 2016. "Blurred lines: Defining social, news, and political posts on Facebook", Journal of Information Technology & Politics, doi: 10.1080/19331681.2016.1160265

[Wang et al. 12] Wang, H., Can, D., Kazemzadeh, A., Bar, F. and Narayanan, S. 2012. "A system for realtime Twitter sentiment analysis of 2012 U.S. presidential election cycle", Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, ACL, 15-120.

[Wen, 14] Wen, W.-C. 2014. "Facebook political communication in Taiwan: 1.0/2.0 messages and election/post-election messages", Chinese Journal of Communication 7(1): 19-39.

[Woolley et al. 10] Woolley, J.K., Limperos, A.M. and Oliver, M.B. 2010. "The 2008 Presidential Election, 2.0: A Content Analysis of User-Generated Political Facebook Groups", Mass Communication and Society 13(5): 631-652.