SOCIAL NETWORK ANALYSIS OF POLITICIAN STATEMENTS ON SOCIAL NETWORKS

Jan Panuš

University of Pardubice, Pardubice, Czech Republic
jan.panus@upce.cz

Abstract

Social networks are an integral part of our lives these days. Most people have an account on some of the many social networks. Social networks are used for various reasons such as to express some opinion or to spreading this opinion. Many subjects take advantage of this situation and one group of such subjects are politicians. This paper aims to analyze the behavior of Czech politicians on social networks. The paper focuses primarily on the statements of politicians on social networks with related characteristics and with specific posts on social networks. The reason why we should be interested in this topic is to control politicians' behavior that affects also the way they express themselves and how they communicate. We used basic statistical tools and some metrics used for social network analysis to find out how this was done. The outcome of the paper expresses mainly the connection between the individual politicians and length of their posts on social networks. The result in the paper also shows the time distribution of individual posts on social networks.

Keywords

Social networks. Social networks analysis. Politicians. Social media.

INTRODUCTION

Social media (SM) represents an integral part of our lives. The individual services offer a great deal of power to control the crowds and convey information to them. Services such as Facebook or Twitter have a wider reach compared to interpersonal communication. SM offers two or more ways to communicate, unlike traditional mass media. SM represents an opportunity for both social interaction and product sales promotion. Twitter might have played important role in various social movements such as protests. A great example is the use of Twitter in the case of the Arab Spring (Khondker, 2011), where using of the network could speed up democratic processes. Recent changes in society give rise to scientific disciplines such as digital mobilization or social change. Not all participants are active or important in these changes of course. They also have a specific involvement in social movements and information transfers however.

The aim of this paper is to introduce possibilities how to identify the most important players in the social networking field using methods for social network analysis and assigning them some common characteristics in order to get an idea of how they can be further analyzed. The paper is focusing on the statements of politicians on social networks. The paper analyzes important metrics such as betweenness centrality, in-degree, out-degree, closeness centrality etc. We analyze each metric and assign each value to a given name. A more detailed description of the research area is describe in the next chapter.

This analysis will help to discover why some individuals are more important and what led them to do so. We can identify significant influencers for spreading information and influencing politician developments. It is possible to make a better prediction of the individuals' behavior and their influence on their social network based on this analysis. An important element is the analysis of the content, which we only touch briefly. A very important part is the analysis of the interaction, which we will not discuss in this article.

THEORETICAL BACKGROUND

Related work

Social Network Analysis (SNA) is used to look for structural relationships or networks that are formed by different organizations or participants (Knoke & Yang, 2019). SNA generally assumes that individual actors are involved in networking by connecting with other actively communicating actors. The behavior of individuals is an important element in establishing further contacts. Such actors could be individuals or social groups, organizations or statuses. The cluster is then a group of elements (actors), where a network of relationships forms further spreads of relationships between other actors. Each network is a living organism, which is constantly changing and creating new relationships. In the case of social media such as Twitter or Facebook, the network element is a specific account, and the cluster is a subset of different accounts linked with each other. It is possible to distinguish individual clusters according to different criteria and then divide them. Each account can create different connections even to unrelated clusters. It is possible to have a few nodes within the cluster as well as several thousands of nodes (accounts). It is generally assumed that individual accounts are communicating within a given cluster rather than outside of the cluster. However, such nodes create an environment for the further dissemination of information throughout the network. Such relationships then have a specific type of connection between nodes (Knoke & Yang, 2019). Relationships can be directed either when interacting or non-directed (mutual friendship). We will talk about the social network account as a node and the connection to another person mentioned below as an edge.

The position of the node within the network can then determine the influence that the node has on the network itself (Burt, Kilduff, & Tasselli, 2013). Influence can be characterized as the effect that someone or something has on the way someone else or something else works or develops (Monge, Peter, Contractor, Contractor, & Noshir, 2003). Nodes can focus on strengthening relationships or building bridges between other nodes or between other clusters (Burt et al., 2013). It is believed that the strength of interaction within a cluster is usually stronger than a connection that is outside the cluster. Strengthening the relationship between the nodes then leads to an improvement in the overall cluster, but it can also have a negative impact that the node can often receive redundant information. On the other hand, if a node is outside the cluster, it can lead to better dissemination of information, because a node that has links to other clusters has the advantage of receiving various information that may not exist within the cluster. All of this, of course, compared to other nodes within the cluster. In this way, an individual who has a weak connection within the cluster but has a connection outside the cluster can affect events in two clusters.

There are a number of centralities in the literature to analyze the behavior of individuals within a social network (Knoke & Yang, 2019; Newman, 2004a, 2004b). In-degree centrality (Hajian & White, 2011) serves to inform about the number of directional links to the actor from other actors whereas out-degree centrality refers to the number of directional links from the actor to other actors.

Betweenness index (Carrington, Scott, & Wasserman, 2005) is another important measure of node's influence in the network. This index means shortest path between pairs of vertices in a network that pass through a vertex (Newman, 2004b). The higher the value is the more influence a node has with spreading information in the network. Such a node is called a broker (Freeman, 1978) or a bridge between different subgroups (Gould & Fernandez, 1989). Communication between subgroups may be very difficult due to different specialization of such groups. Betweennes centrality is more important metric compared to degree centrality.

Closeness centrality is another important metric that measures average distance to all other nodes in the network. The higher score is the shortest distance is to all other nodes. Individuals with high closeness centrality are in position to control and acquire information within the organization (Krebs, 2002). Closeness centrality is possible to interpreted as an estimated time of arrival of information within telecommunication or package delivery networks (Borgatti, 2005).

This paper should bring new view on social media communication between politicians in Czech republic. Paper combine different statistics for measuring activities on social media.

EMPIRICAL RESEARCH

Method

For the purposes of this work, we have analyzed the statements of politicians on Czech social networks, specifically on Twitter and Facebook. The data comes from a database of politicians' statements (Bláha, 2020), which is constantly updated. Due to the large amount of data, we have analyzed only the data available for the entire year 2019. The database collects individual statements (or posts) of all politicians who have an account on social network. The database contains 172 819 records for the entire period which is from 2009 to the present days. The activity of politicians was not at the same level at the beginning as it is in these years. We selected such records where individual politicians refer to another politician for the purpose of this paper. We conducted basic statistical analysis and social network analysis based on this data. The total number of these records is 88 297. Therefore, analysis performed only for records that were conducted in 2019 and the number is 28 858.

We created a JSON file from these data and we analyzed primarily using applications that are suitable for social network analysis. We used C # language in the Visual Studio development environment and subsequent analysis was performed in NodeXL application that processes data in MS Excel. Based on this data, we then generated a network of elements that depicted specific links to each politician. Due to the large number of both politicians and references, we have made a reduction, which is particularly evident in the following tables. We have always selected only the most active politicians to show individual results. In some cases, there is only one politician reference, and a graph that depicts such a situation would be very confusing. All data and descriptions are then based on the default database (Bláha, 2020), so as not to disturb the overall order of the data. Therefore, we kept the names of politicians in the format provided by the database. Another essential part of the analysis of the politicians' statements is the analysis of the textual part. We show basic statistics in this paper and we discuss this statistic in the next chapter.

Model testing results

Table 1 shows the total number of words that each politician wrote for the year of 2019 on both social networks that we followed. As you can see, the most active politician is Tomio Okamura, with 649 889 words, with the greater part occurring on Facebook (more than 95% of all written words). In second place is Prime Minister Andrej Babiš with a total of 251 111 words, with Facebook posted over 83% of all words. In any case, the total number of words is only 38,6% of what Tomio Okamura wrote. Other places are followed by Petr Fiala (13,8% of total words against Tomio Okamura), Miroslav Kalousek (9,5% of total words against Tomio Okamura) etc.

From the overall overview, it is clear that most politicians are more active on Facebook rather than Twiter. One important reason is that only 140 characters it is possible to write to one post on Twitter, while the length of text on Facebook is virtually unlimited. However, one of the exceptions is Miroslav Kalousek, who is more active on Twitter. This may also be because the target group of voters appears more on Twitter where Kalousek is very active.

Table 1: Number of words for each of social media

Politician	Count of words		
	Facebook	Twitter	Total
tomio-okamura	620020	29869	649889
andrej-babis	208548	42563	251111
petr-fiala	64593	25694	90287
miroslav-kalousek	15066	47023	62089
alena-schillerova	37068	23400	60468
vladimir-kremlik	39570	13993	53563
jan-zahradil	14374	37475	51849
jana-vildumetzova	46640	935	47575
adam-vojtech	31071	15439	46510
jan-hamacek	20503	21315	41818
karel-havlicek	7445	33089	40534
zdenek-hrib	29807	5201	35008
radek-vondracek	21595	11174	32769
jan-bartosek	18365	13116	31481
alexandra-udzenija	22933	8181	31114
vera-jourova	0	30791	30791

Table 2 shows the average number of words per post. Word count affects how a particular post is perceived (Yoon, Syn, & Tippett, 2019) and it is definitely not appropriate to have very long posts. However, there are politicians who do not perceive this and have long contributions. A typical example is Tomio Okamura with an average of 231,85 words per post. This is followed by Petr Vokřál with 155,53 words per post (approximately 67% of what Okamura), Jaroslava Jermanová (about 59% of what Okamura) etc. Andrej Babiš has very long posts, but with an average of 79,54 words per post occurs only in 10th place.

Table 2: Average number of words per post

Politician	Count of words		
	Facebook	Twitter	Total
tomio-okamura	335,87	31,21	231,85
petr-vokral	158,21	43,00	155,53
jaroslava-jermanova	138,48	0,00	138,48
radek-holomcik	147,84	16,00	119,30
jana-vildumetzova	127,08	29,22	119,24
jana-pastuchova	113,11	0,00	113,11
milos-zeman	103,35	0,00	103,35
ondrej-kolar	101,56	0,00	101,56
klara-dostalova	99,49	0,00	99,49
andrej-babis	134,11	26,57	79,54
vladimir-kremlik	133,68	33,48	75,02
herbert-pavera	74,82	36,31	69,86
zdenek-hrib	93,15	26,67	67,98
jan-farsky	65,51	0,00	65,51
vlastimil-valek	118,85	28,37	62,01
petr-hladik-11	80,30	25,42	61,46

Figure 1 shows the total number of words per month for the five selected politicians. They are mainly the most active. As is evident, for example, Andrej Babiš was not so active on Facebook throughout the year, but the total number of his comments rose during September. This activity may have a reason in a published case with subsidies from Agrofert. Tomio Okamura has been more or less active throughout the year, but he also rose significantly at the end of the year. Other politicians are no longer so active. Data for December is low, which may be because Christmas season when activity of politicians on social networks is declining.

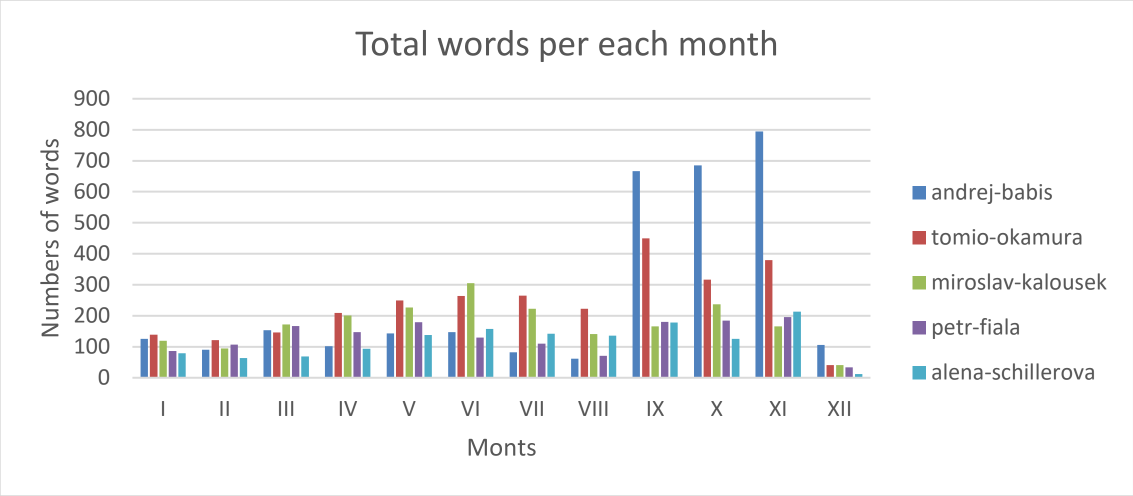

Figure 1: Total words per each month for five most active politicians

An important element is the analysis of connections with other nodes within the network. The following table 3 is sorted by in-degree centrality, which expresses how many other politicians referred to that politician. Table shows only chosen politics, total number with this centrality would be 36. Prime Minister Andrej Babiš has the higher score of in-degree centrality. Results show that politician do not use the opportunity to connect to other politicians much. However, what the use is an indicator of out-degree centrality - how many other politicians refer to specific politician. Tomio Okamura is very active, he is referring to up to 255 different politicians. Andrej Babiš is on second place.

The highest score of betweennes centrality has Tomio Okamura again and Andrej Babiš is on second place. High value of Okamura's score is given by his activity not by activity of other politicians. The reason is his very low in-degree centrality score. The high value of Andrej Babiš is given by the fact that he is the prime minister and he is also part of problematic cases at the same time. Closeness centrality scores are very low generally, which is mainly because individual politicians are not closely connected to each other and because there are a large number of politicians in the network who refer to only one other politician.

Table 3: Metrics for chosen politician

Politician	In-degree	Out-degree	Betweenness centrality	Closeness centrality
andrej-babis	38	153	103242,8293	0,00105
jan-hamacek	17	41	14540,34118	0,000864
alena-schillerova	15	26	9866,758064	0,000833
adam-vojtech-1	14	19	6340,65711	0,00083
vera-jourova	8	2	935,222259	0,000786
petr-fiala	8	67	16936,43958	0,000847
tomio-okamura	8	255	143845,9421	0,001041
jan-zahradil	7	34	11394,53385	0,000812

radek-vondracek	4	22	5319,925898	0,000743
alexandra-udzenija	0	17	796,290424	0,000708
jana-vildumetzova	0	26	8553,50389	0,000696
karel-havlicek-30	0	10	68,237207	0,000678
vladimir-kremlik	0	25	7250,926622	0,000702
zdenek-hrib	0	34	6272,728734	0,000709

Figure 2 illustrates the visualization of communication between politicians. Due to the large amount of data, it was necessary to make some reduction. Politicians have been selected with an in-degree centrality higher than eight. Arrows represent directionality in the graph. The strength of the link then expresses the value of in-degree centrality, the stronger the link is the greater the in-degree value. It is possible to create many similar graphs but we chose just this such an illustration of better orientation in the given network.

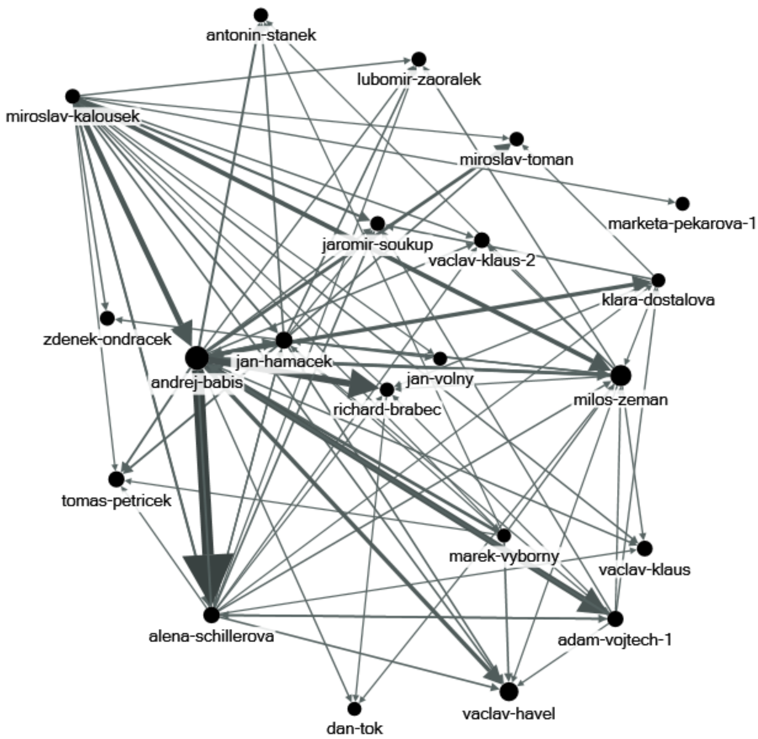

Figure 2: Visualization of communication between politics

DISCUSSION

We conducted an analysis of the communication of Czech politicians on social networks in this paper. Facebook and Twitter were chosen as social network because these networks are widely used in Czech Republic. The data were obtained for the year of 2019 and include individual posts on social networks, date of origin, number of words in each post and links to other politician. We performed basic statistics analysis and subsequently analysis of behavior on social networks. It is recognizable which politicians are active and how they express themselves on social networks. Obtained data show mainly the number of words in one post and in what period politicians are active. The reference to other politicians is also an important aspect. However, they do not use much of this feature of social networks.

The data show that a very active member is the Prime Minister. It is known that his account manages more people. Tomio Okamura is also very active on social networks. The Prime Minister is also an active element in referring to other politicians. This fact may be due to ongoing cases around Andrej Babiš.

Betweenness centrality means the shortest path between every pair of vertices in connected graph. The betweenness centrality for each vertex is the number of shortest paths that pass through the vertex. The higher value is the more important the vertex is in meaning of connecting some subgroups. High value of the betweenness centrality has Tomio Okamura and Andrej Babis.

Closeness centrality measures the mean distance from one vertex to another vertex and means shortest path through a network between two vertices. The lower value is the better access information at other vertices is. Actually, the closeness centrality is very low in our study. It means that politician does not have strong connection between each other.

A very interesting element of further analyzes would certainly be the content analysis, especially with regard to finding the topic of the text with further analysis designed to determine the temporal and spatial aspects. It would be very interesting to find out whether there are certain patterns in the written text and whether these patterns are using to deduce who the politician is. Another interesting way could be to find out the mood in the context with corresponding mood of politician's electoral base.

We selected data just for year of 2019 but previous period data are available also therefore it would be interesting to find out how the relationship between politicians evolved through the time. It is possible to deduce politicians' behavior from data on current political cases but we should verify this statement. A typical example is the increased activity of the Prime Minister of the Czech Republic in the period of September - November, when an effort to transfer attention to other topics can be expected. This would be suitable for text and content analysis. In the future, it is assumed that I will focus more on these topics.

CONCLUSION

The data collected and analyzed provide a better picture of the overall society and the behavior of politicians on social networks. The analysis confirms the validity of both the data obtained and the analysis itself, which is appropriate to understand how individual politicians create their image outwardly. The analysis also confirm that politicians should use social networks for their promotion and better control of their behavior. If the textual and content analysis were also carried out, it would be useful to review them and draw attention to what they have published in the past. Politicians also mostly focus on Facebook, which is also because this social network is the majority of the population of the Czech Republic. However, it is also possible to focus primarily on other networks. However, this would be for further analysis, especially for the behavior of the electoral base of individual politicians. In the future, I would like to focus primarily on the content analysis of individual messages and on the comparison of behavior over time.

REFERENCES

Bláha, M. (2020). Databáze Vyjádření politiků. Retrieved 13.2.2020, from Hlídač Státu z.ú https://www.hlidacstatu.cz/data/Index/vyjadreni-politiku

Borgatti, S. P. (2005). Centrality and network flow. *Social networks, 27*(1), 55-71.

Burt, R. S., Kilduff, M., & Tasselli, S. (2013). Social network analysis: Foundations and frontiers on advantage. *Annual review of psychology, 64*, 527-547.

Carrington, P. J., Scott, J., & Wasserman, S. (2005). *Models and methods in social network analysis* (Vol. 28): Cambridge university press.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social networks, 1*(3), 215-239.

Gould, R. V., & Fernandez, R. M. (1989). Structures of mediation: A formal approach to brokerage in transaction networks. *Sociological methodology*, 89-126.

Hajian, B., & White, T. (2011). *Modelling influence in a social network: Metrics and evaluation.* Paper presented at the 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing.

Khondker, H. H. (2011). Role of the new media in the Arab Spring. *Globalizations, 8*(5), 675-679.

Knoke, D., & Yang, S. (2019). *Social network analysis* (Vol. 154): SAGE Publications, Incorporated.

Krebs, V. E. (2002). Mapping networks of terrorist cells. *Connections*, *24*(3), 43-52.

Monge, P. R., Peter, R., Contractor, N. S., Contractor, P. S., & Noshir, S. (2003). *Theories of communication networks*: Oxford University Press, USA.

Newman, M. E. (2004a). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the national academy of sciences, 101*(suppl 1), 5200-5205.

Newman, M. E. (2004b). Detecting community structure in networks. *The European Physical Journal B, 38*(2), 321-330.

Yoon, J., Syn, S. Y., & Tippett, A. (2019). How do people perceive Facebook features on health-related Facebook posts? *iConference 2019 Proceedings*.