IBIMA Publishing
Communications of the IBIMA
https://ibimapublishing.com/articles/CIBIMA/2021/589003/
Vol. 2021 (2021), Article ID 589003, 18 pages, ISSN: 1943-7765
DOI: 10.5171/2021.589003

Research Article

"An Effective Multidimensional Model for Analyzing Social Web Big Data – Testing in simple Web 2.0 Applications of Internet Politics"

Dimitrios VAGIANOS[1] & Kostas ZAFIROPOULOS[2]

[1]Electrical & Computer Engineer, MSc, PhD - Laboratory Teaching staff, Department of International and European Studies, University of Macedonia, Egnatia, Thessaloniki, Greece

[2] Department of International and European Studies, University of Macedonia, Egnatia, Thessaloniki, Greece

Received date: 10 January 2019; Accepted date: 20 November 2019; Published date: 23 February 2021

Abstract

Web 2.0 applications have provided researchers with vast quantities of Big Data, opening up new horizons for developing innovative analysis techniques that are applicable in multiple cognitive fields. Politics is definitely among them and is currently in a dynamically evolving range of these applications. Although modern Social Networking Platforms dominate the digital political landscape by supporting impressive volumes of political Big Data, political trends could nevertheless be traced in the pioneering network of one of the first participatory web applications: the blogs. The blogs and their social network, the so-called blogosphere, featured most of the advanced characteristics of contemporary Social Media Platforms, in a more simplistic form. After 15 years of digital existence, they maintained a noticeable presence in the Social Web while they have additionally undertaken the role of gateways to the multifaceted realm of Social Media Networks. Since their introduction, they have been providing digital citizens with a user-friendly tool to post political content, mutually interact, shape the political agenda and horizontally influence the public opinion as well as vertically affect the administrative decision centers. Therefore, they can be used as a reach Big Data fields suitable for testing multidimensional modeling methods towards mapping political trends, which could be upgraded accordingly for use in more complicated Social Media Applications. In this paper, a qualitative as well as quantitative method for analyzing the political blogosphere is introduced, consisting of three components. It focuses on the formation of blog communities, based on their hyperlink interconnectivity, the blogs' influence and their users' generated content. By applying a Multidimensional Scaling and Cluster Analysis, clusters were located, which correspond to the afore mentioned communities. These clusters, were found to be somehow related to the degree they may potentially influence their readers. Findings showed that high influence does not always involve high rates of hyperlinking. Eventually, by applying Content Analysis to the content of the blogs forming the clusters, the qualitative characteristics and the general topics of the political discussion were identified, regarding the specific survey period. Findings suggested that the qualitative characteristics of the political blogs are significantly related to the cluster they belong to, according to their hyperlinking. In the case study presented, the method proved to be able to investigate both the political qualitative and quantitative blogosphere characteristics. By applying it in the Greek political blogosphere, it proved to be able to thoroughly investigate the political debate that takes place on a portion of the social web, its characteristics and the influence it potentially has on public opinion.

Keywords: Political Blogosphere Big Data, hyperlink Analysis, Influence Analysis, Content Analysis.

Cite this Article as: Dimitrios VAGIANOS & Kostas ZAFIROPOULOS (2021)," "An Effective Multidimensional Model for Analyzing Social Web Big Data – Testing in simple Web 2.0 Applications of Internet Politics"", Communications of the IBIMA, Vol. 2021 (2021), Article ID 589003, DOI: 10.5171/2021.589003

Introduction

Blogs, as the first representatives of Web 2.0 applications, have significantly contributed to shaping the digital political realm across the globe, mainly due to their simplicity. Therefore, they have been envisaged as a means of political expression that had the power to shape political agendas and political dialogue (Wallsten, 2005). They have also been considered as reliable alternative news sources capable of spreading information and mobilizing citizens incredibly quickly (Albrecht et al., 2007). Howard Dean in 2004 was the first politician that used his blog exclusively for supporting his political campaign, being the first candidate to raise a significant amount of money exclusively derived from his blog campaigning activity (Graf, 2006). His methods inspired many succeeding candidates in the United States with George Bush Jr among them. The US 2004 elections were the key fact that demonstrated the key role and influence of political blogs (Keren, 2006). In the political blogosphere all individual diverse contributions with posts and comments are highlighted, underlying the power of each individual contributor, which is in full opposition to the role citizens have in the Mass Media landscape (Reynolds, 2006). Due to these factors, the number of political blogs has been ever increasing in the years that followed, even though several other means of Social Media applications with political involvement have been gradually introduced. One of the blogs' main characteristics is their interlinking capability in forming their own social networks (Du & Wagner 2006). Therefore, visitors can easily perform blog hopping following hyperlinks, discovering new sources of information, interacting with strangers but most importantly forming patterns of productive collaboration (Efimova & Hendrick, 2005).

According to Jackson (2006), political blogs can exhibit an asymmetrically high influence based on the characteristics of their visitors and not their number of visitors. Therefore, prominent political blogs can play a pivotal role in attracting visitors exerting an influence on public opinion. Because of this asymmetric distribution, there are ultimately few discrete political blogs that play a key role in the accumulation and dissemination of information and all of these blogs may eventually contain a summary of the qualitative characteristics of the entire blogosphere (Drezner & Farrell, 2004).

From what was stated above, it is evitable that the political blogosphere soon became a rich tank of Social Web Big Data that can be exploited for a variety of purposes: discovering patterns, finding influential entities and discovering main topics of interest. In this paper, a three-dimensional analysis method is introduced that aims at providing this information and generally mapping the political blogosphere in order to extract useful results that can be politically exploited in a variety of political aspects of Social Web applications.

Related Research History

Adamic and Glance in 2005 measured the degree of interaction between liberal and conservative blogs in order to uncover differences in the structure of the two communities. They based their research on blogrolls in order to present a static picture of the political blogosphere in the US. Their outcomes were based on a single day snapshot of over than 1,000 political blogs. Du & Wagner (2006) underlined that the blogroll hyperlinks give a societal identity to blogs and give an interactive dimension to their sphere (Albrecht et al., 2007). Visitors' participation in terms of comments can contribute to social relationship development within the political blogosphere (Ali-Hasan & Adamic, 2007). Sigala (2008) noted that an analysis of blogs' social network patterns can potentially lead to communities' identification while many research projects started investigation over potential influence of the blogs towards shaping the daily agenda and public opinion (Wallsten, 2007).

Zafiropoulos and colleagues (2011) proposed a statistical method of analyzing patterns in Social Network of blogs of specialized thematic.

Table 1: Published blog impact assessment studies based on data from specific sources

Researcher's name and corresponding work	Analysis type	Origin of data
McKenna & Pole (2004), "Do Blogs Matter? Weblogs in American Politics"	Study of a blogs' sample based on ranking lists	Blogstreet Truth Laid Bear Ecosystem Technorati Truth Laid Bear EcoTraffic
Drezner & Farrell (2004), *"The Power and Politics of Blogs"*	A detailed overview of the political blogosphere	Technorati Truth Laid Bear Ecosystem Blogstreet
Gill (2004), *"How Can We Measure the Influence of the Blogosphere?*	An overview of influencing monitoring mechanisms	Blogosphere.us Blogrunner Blogstreet Technorati
Hindman (2005), *"Voice, Equality and The Internet"* (doctoral dissertation)	Bloggers' social survey based on online blogs' ranking	Truth Laid Bear EcoTraffic
Adamic & Glance (2005), *"The Political Blogosphere and the 2004 US Election: Divided they Blog"*	Hyperlinks' blog investigation in the political blogosphere	Blogpulse, with comparisonsto Technorati Truth Laid Bear EcoTraffic Truth Laid Bear Ecosystem
Auckland (2005), *"Mapping the U.S. Political Blogosphere: Are Conservative Bloggers More Prominent?"*	Network analysis of the A-listed blogs with Uberlink software	Adamic & Glance
Wallsten (2007), *"Political Blogs: Transmission Belts, Soapboxes, Mobilizers or Conversation Starters."* (doctoral dissertation chapter)	Study of a blogs' sample based on ranking lists	Blogstreet Truth Laid Bear Ecosystem Technorati Truth Laid Bear EcoTraffic

Table 1 shows several past research attempts that aimed at measuring the impact of blogs by using specific datasets and software.

All research attempts investigated parts of the blogosphere from a single perspective, leaving aside important aspects of the evolving phenomenon. In most cases, the datasets under investigation were not particularly large and did not take into consideration the quick proliferation of Big Data in the blogosphere available for mining. Other Social Media Platforms such as Facebook and Twitter soon became the most popular tools for users' interaction and spreading of User Generated Content.

Inevitably, research was directed towards investigation over patterns of these innovative social networks. However, after 15 years of digital existence, blogs have always been a constant gateway to the multifaceted realm of Social Media Networks. Bloggers around the world still locate their target groups within the blogosphere to diffuse their digital content, but now, advanced Social Media platforms are at their disposal to boost this process. Blog interlinking has always been a simple method to support this diffusion process although now the situation is more complicated (Vagianos, 2019).

This paper introduces a multifaceted method that embodies all important components in a complementary logic which can be used and flexibly further developed by researchers within the timeless blogosphere.

Model's Methodology

The first step in the proposed methodology was to define a representative sample of the

political blogosphere which adequately reflects the properties of the total number of political blogs. A second sampling should be done by defining a specific time window in which the blogs' data should be retrieved.

The proposed multidimensional model that has been used for the Analysis of the representative sample of blogs included the following: Cluster Analysis over the social network in the sample of blogs, Influence Index Analysis and Content Analysis over the posts of the selected blogs of the sample during the selected time window. The results of each step should be compared with the results of the other two leading to the final properties of the representative groups of political blogs under investigation.

The first step of the method is the statistical part. After defining the blogs' sample, an adjacency matrix is formed containing binary digits indicating if there is a hyperlink connection between two blogs or not. The square non-symmetric adjacency matrix containing ones and zeros is forming a multidimensional space whose dimensions should be reduced. This is accomplished by applying the Multidimensional Scaling technique. Subsequently, The Two-Step Cluster Analysis technique is applied for identifying cluster formation in the total social network of the blogs under investigation. This first step provides an initial overview of how the bloggers community considers the blogs' network in terms of how these blogs congregate in groups in terms of their inbound hyperlinks. Additionally, this step aims at identifying the A-list blogs (Trammel & Keshelashvili, 2005). According to Drezner and Farrell (2004), the A-list blogs embody a "Statistical Summary" of the mainstream of the view of the overall blogosphere regarding a certain political issue.

The second step focuses on the Influence exerted by each blog. Following the approach of Karpf (2008), four measures of influence were used: the Network Centrality Score, the Hyperlink Authority Score, the Site Traffic Score and the Community Activity Score. The measurements of the four indexes for each blog are transformed into ranks. Following Karpf (2008), the total Blogosphere Authority Index is the sum of the ranks of the four measures minus the worst rank of them. Having calculated all the four indexes along with the total Blogosphere Authority Index, 5 independent listings are formed that describe the influence of each blog from a different perspective. These results can be envisaged within both the resulting clusters of the first step of the method, validating cluster formations or giving a better overview of groups of blogs found not to belong to clusters regarding common incoming hyperlinks but still having enormous influence due to being visited by a large number of readers. They can also be viewed from an independent perspective providing listings in the total sample of blogs under investigation, no matter to which cluster they were found to belong in the first step of the method.

The third step involves a Data Mining technique. More specifically, it introduces Content Analysis over the posts of the selected blogs in the selected time window. Although the incoming hyperlink analysis of step 1 and comment counting (calculation of Community Activity Index) of step 2 fall within the broad context of Big Data analysis (Herring et al. 2009), a deeper analysis technique has been introduced in order to embody qualitative research in the overall technique. The results are both compared and combined with those of steps 1 and 2 providing an integrated overview of the popular topics of interest in the bloggers' community in accordance with the clustering formation as the exerted influence of each blog.

Each component of the proposed method was complementing the other two. As it will be shown in the next chapter, parts of the proposed method can be automatically implemented where others require a fully manual approach. Although some assumptions should unavoidably be made

during the procedure, the final result achieves a thorough overview and a detailed mapping of the so called Activated Public Opinion as it is expressed in the political blogosphere.

Model Testing in Greek political blogosphere

The proposed method has been implemented and tested over the Greek political blogosphere. Political blogs started to exert a significant influence after 2008 (Zafiropoulos & Vrana, 2009) and by 2011 have been considered to possess their own merit in the Greek political realm forming the Greek political blogosphere which would play its own role under the prevailing local debt crisis circumstances and the adopted policies landscape and the forthcoming elections that would displace the two prevailing political parties and bring the left party "SYRIZA" into power in year 2015. Therefore, the first two months of year 2011 have been selected as the time window in which the testing of the method would be deployed.

Component 1: Multidimensional Scaling and Cluster Analysis

During this period, the appropriate key words (names of all the Greek political parties) have been applied to Google blog Search Engine (http://www.google.com/blogsearch) as well as Technorati.com in order to record the most popular political blogs. In this approach, it was the Search Engines' ranking algorithm (Brin & Page, 2008) that formed the listingsaccording to the page popularity. The blogroll of each blog has not been used to lead to new blogs in an attempt to eliminate biasing of the results when the incoming hyperlinks of each blog would be analyzed. During this procedure, 127 Greek political blogs have been recorded, which would form the representative sample of the Greek Political blogosphere. As a next step, Website Extractor (10.2 version Internetsoft Corporation) has been used as a tool to retrieve the html code of each blog while Microsoft Visual Basic 6.0. has been used to create a script that would scan the blogs' html codes and create the 127x127 square non-symmetric adjacency matrix containing ones or zeros in cell *ij* depending on an incoming hyperlink existence or non-existence of the blogroll of the blog in row *i* to the blog of column *j*.

Since the automatically produced adjacency matrix describes a 127-dimensional space, it was practically important to ideally reduce it to a two-dimensional one. This has been accomplished (figure 1) by applying the Multidimensional Scaling technique by using the SPSS software package (version 20) and its PROXSCAL routine. The log-likelihood distance has been selected as it led to a better (<0,1) *Stress* Index (Kruskal, 1978), which in this implementation has been found equal to 0,0453.

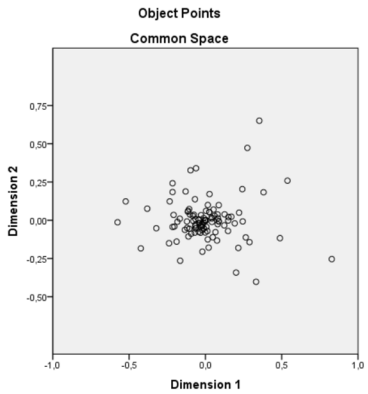

Figure 1: Projection of the 127 blogs on the two-dimensional space

In order to identify possible clustering, the two-step Cluster Analysis method has been applied. In this procedure, the Bayesian information criterion (BIC) has been selected as a way to avoid overestimating the final number of clusters (Burnham & Anderson, 2004). The resulted number of clusters was 4 in this case study, considering the SPSS' criteria regarding cohesion and separation (figure 2).

Figure 2: SPSS 20 two-step Cluster Analysis result

As it is shown in table 2 the BIC value for 4 clusters has been found to be 157,538.

Table 2: Two-step Cluster analysis and automatic calculation of optimal number of clusters according to BIC

Quantity of Clusters	BIC (Schwarz's Bayesian Criterion)	BIC* gradual Difference	BIC** Difference Ratio
1	194,434		
2	168,756	-25,678	1,000
3	159,832	-8,924	,348
4	**157,538**	**-2,294**	**,089**
5	162,872	5,334	-,208
6	175,260	12,388	-,482
7	188,772	13,512	-,526
8	202,790	14,018	-,546
9	216,888	14,098	-,549
10	231,103	14,214	-,554
11	245,680	14,577	-,568
12	261,778	16,098	-,627

13	278,404	16,626	-,647
14	295,092	16,689	-,650
15	312,264	17,172	-,669

Cluster 1 contained 21 blogs, Cluster 2 16 blogs, Cluster 3 was found to be formed by 18 blogs and the remaining 72 blogs were found to form the 4th Cluster. The number of the incoming hyperlinks of the blogs of each cluster is shown in table 3.

Table 3: Average of incoming hyperlinks of blogs of the 4 identified clusters

	Average of incoming hyperlinks (standard deviation)	**Average of the percentage of the incoming hyperlinks (standard deviation)**
Cluster 1 (21 blogs)	5,81 (2,23)	4,57 (1,75)
Cluster 2 (16 blogs)	7,26 (2,85)	5,80 (2,24)
Cluster 3 (18 blogs)	11,33 (5,40)	8,92 (4,25)
Cluster 4 (remaining 72 blogs)	1,69 (1,44)	1,33 (1,13)
Sum	4,46 (4,40)	3,51 (3,46)

At a glance, it can be noticed that Cluster Analysis recognized 3 clusters containing a total of 55 blogs, based on the way they are evaluated by the bloggers' community. According to Herring et al. (2004), clusters 1,2,3 and can be viewed as the central ones in the blogs' social network formed by their incoming hyperlinks. Figure 3 shows the 4 clusters and the predictive importance of each one of the d1, d2 variables of the two-dimensional space.

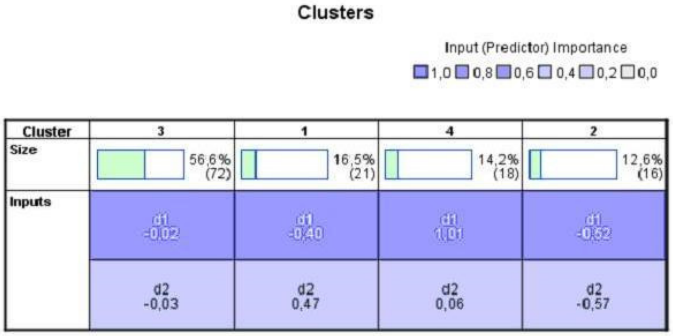

Figure 3: Clusters' size and predictive importance of each one of the d1, d2 variables of the two-dimensional space

By using Ucinet for Windows (version 6, Harvard Analytic Technologies 2002) the 3 central clusters' network are shown in figure 4 while the 72 remaining blogs of cluster 4 are shown as a "blog cloud" around them.

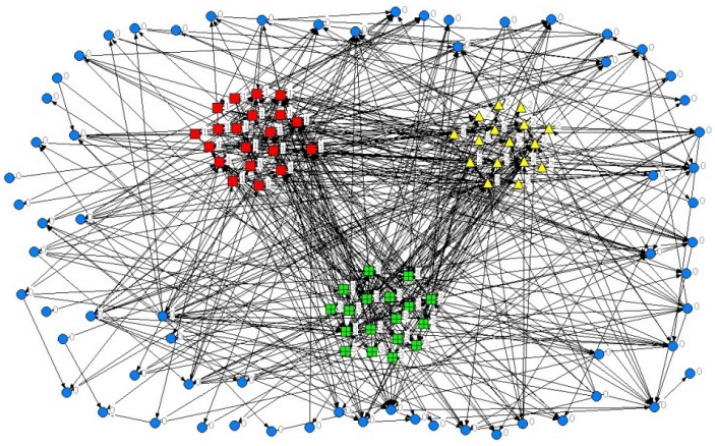

Figure 4: Blogs' Social Network graph with 3 clusters (21,16 and 18 blogs) in the middle and the rest 72 blogs forming a blog cloud around them.

Commenting on the results at this point, Clusters 1, 2 and 3 are highly probable to include the A-list blogs mentioned above while Cluster 3 is highly probable to be formed by the most popular and therefore most influential blogs, given that visitors are led to the blogs of this cluster instructed by the blogrolls of the rest of the blogs' sample.

It must also be noted that from a top level Content identification, Cluster 1 contained blogs belonging to the left parties: mainly SYRIZA and KKE.

Finally, the degree of connection between clusters is shown in table 4.

Table 4: Clusters' degree of Connection

Clusters degree of connections (%)				
	Cluster 1	**Cluster 2**	**Cluster 3**	**Rest of Blogs**
Cluster 1	76,19	42,86	23,81	38,09
Cluster 2	43,75	56,25	18,75	62,5
Cluster 3	38,89	44,44	94,44	61,11
Rest of Blogs	25	27,72	33,33	36,11

The Degree of Connection between two clusters is defined as the percentage of the blogs of a cluster that has at least one hyperlink towards another cluster. The same logic can be also adopted when dealing with hyperlinks within the same cluster. Table 2 shows that Cluster 3 has the highest internal cohesion with 94,44 of its blogs redirecting visitors traffic towards a blog of the same cluster.

Component 2: Influence Analysis

The analysis carried out in step one focused on the way the elite bloggers' community considers the blogs themselves and on the

redirection of the traffic between them following blogrolls' hyperlinks. Although there is clear evidence that some blogs, especially the ones forming Cluster 3, dominate the game in terms of receiving large numbers of visitors, it hasn't been quantitatively proved that they exert influence to the bloggers community as well as the blogosphere visitors as a whole. The second component of the proposed method in this paper introduces a way to analyze this aspect by complementing and interpreting findings of step 1.

As stated in the methodology part, the approach of Karpf (2008) is adopted. Four measures of influence were used: the Network Centrality Score, the Hyperlink Authority Score, the Site Traffic Score and the Community Activity Score. This approach is taking into consideration four aspects of visitors' activity in the political blogosphere in an attempt to finally form an integrated Blogosphere Authority index that adequately describes the overall influence of a blog independently from the ranking alogorithms of the Search Engines.

The Network centrality index is calculated as the normalized betweenness of each blog regarded as a node within the blogs' social network. Blogs are the nodes while incoming links are the connections in the network. Loosely, betweenness equals the number of geodesic paths that pass through a node or the number of "times" that a node needs a given node to reach any node by the shortest path. A blog being between many probable geodesic paths increases its probability of attracting visitors which may have an effect on its influence on the community (Hanneman & Riddle, 2005). Normalized betweenness divides simple betweenness by its maximum value. Normalized betweenness was calculated using Freeman's approach with UCINET 6 for Windows (Borgatti et al. 2002).

Hyperlink Authority Index has been usually measured using Technorati.com authority index. However, since Technorati.com cannot provide such indexes for non-English language blogs, Sync.gr popularity index was used instead. Sync.gr was founded in January 2007 and has been the biggest blog aggregator in Greece. On the 3rd of February 2011, sync was identified through Alexa.com having an Alexa Traffic Rank 27698. Traffic Rank in Greece was 180 and 656 Sites were Linking In it. Sync.gr introduced itself as "a meeting point for anything written in Greek blogosphere". Though, Sync.gr users can register their blogs, upload podcasts, videos and photos; use advanced features for showing updates from other social networks; connect with friends and track their notifications.

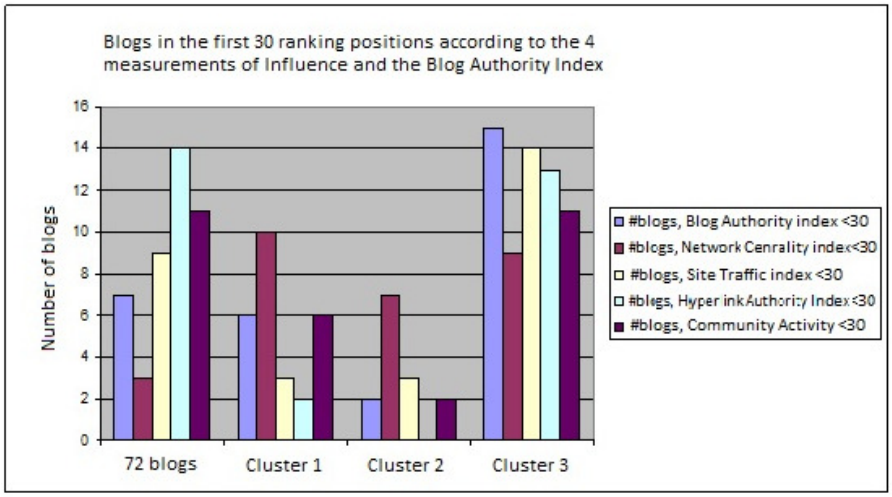

Figure 5: Number of blogs in the first 30 ranking positions according to all side influence indexes and according to the overall Blog Authority Index.

Site Traffic index Score was recorded using Alexa.com and finally Community Activity Score was recorded for each blog in the analysis using a portion of big data of the blogosphere. It measured the number of comments to the posts of the blogs during one week frame, which in this case study was set to the period between 24th of January and 30th of January 2011. Big numbers of comments correspond to a big community involved in specific topics of discussion and therefore the potential to influence the blog's visitors.

The measurements of the four indexes for each blog were transformed to ranks which resulted in 4 different rankings of the 127 representative blogs. These rankings, combined with the clustering results of step 1, led to a clear image of the congregation of the blogs according to their hyperlinks. Figure 5 shows the number of blogs in the first 30 ranking positions according to all side influence indexes.

The Blogosphere Authority Index is the sum of the ranks of the four measures minus the worst rank of them. As Karpf (2008) notes, this is to avoid unfairly biasing the study against sites whose architecture does not allow for reader comments and to minimize outlier effects that come from flaws within any of the four measures employed. The final ranking equation (Karpf, 2008) is:

$$Rankfinal=Rank1 + Rank2 + Rank3 + Rank4 - WorstRank \quad (1)$$

According to (1) the best possible score is 3, indicating that a blog was first-ranked in three categories. The worst possible score is three times 127= 381, indicating that a blog was last-ranked in all categories. In the present study, the values of this final index are also ranked, in order to be easy to comprehend. Figure 4 shows the number of blogs in the first 30 ranking positions according to the overall Blog Authority Index as well.

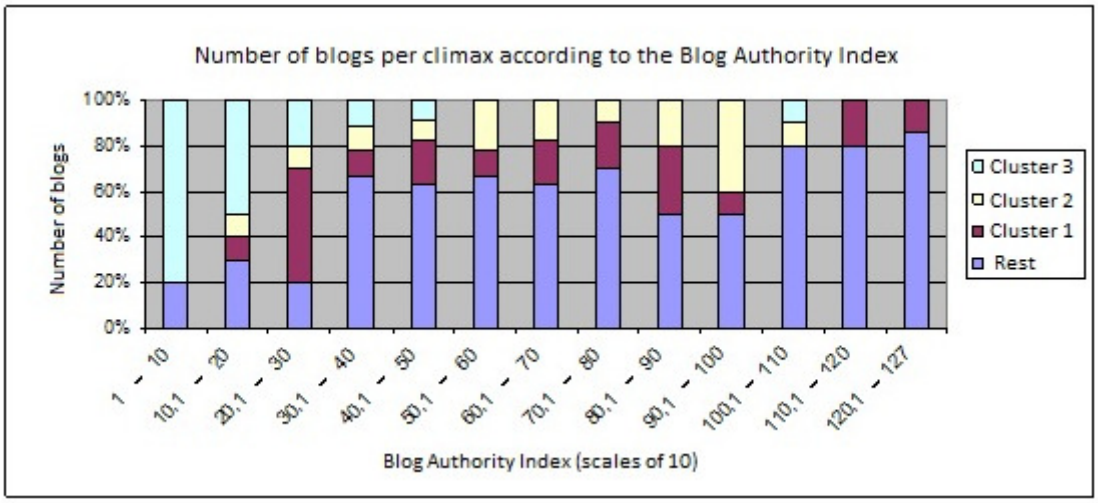

Figure 6: Number of blogs according to the overall Blog Authority Index in scales of 10.

As a general comment, it should be noted that a smaller rank means a better placement in the list of any index used in the study. For example, rank 1 is assigned to the blog which is first in the list for a particular index. From another perspective, table 6 shows the number of blogs according to the overall Blog Authority Index in scales of 10.

Before commenting on the results of the Influence analysis, the correlation of the 4 side influence indexes is calculated in an attempt to investigate the exerted influence of the blog which is somehow related to the blogs' position in their social network. The correlation was calculated based on the Pearson correlation coefficients shown on table 5. It can be seen that the normalized betweenness is not correlated with the community activity and the site traffic index, while the incoming hyperlinks are correlated with the normalized betweenness and the community activity. The incoming hyperlinks are partially correlated with the site traffic. The basic result of the correlation check is that the way the influence indexes were defined does not imply any correlation between them which means that there is no overlapping hidden information among them and that each index leads to a different ranking based on a totally different perspective of Influence of the 127 blogs.

Table 5: Pearson Correlation coefficients of the 4 Influence indexes of the 127 blogs.

	Incoming links	Normalized betweenness	Alexa traffic rank	Sync rank
Normalized betweenness	0.531**			
Alexa traffic rank	-0.287**	-0.109		
Sync rank	-0.176	-0.003	0.624**	
Comments	0.508**	0.081	-0.146	-0.206*

***: p< 0.01, *: p< 0.05*

In Table 6, the average ranking of the blogs of each cluster is shown, both regarding the side influence indexes and the overall Blog Authority Index.

Table 6: Average Rankings of Clusters according to all Influence Indexes

Cluster (# blogs)	Incoming hyperlinks (standard deviation)	Betweenness (standard deviation)	Authority (standard deviation)	Site Traffic (standard deviation)	Community Activity (standard deviation)	Blog Authority Index (standard deviation)
1 (N=21)	40 (16,78)	40 (31,25)	61 (25,81)	78 (32,39)	66 (36,57)	**63 (33,88)**
2 (N=16)	30 (15,45)	51 (42,71)	70 (24,21)	67 (28,35)	73 (32,62)	**67 (27,74)**
3 (N=18)	16 (12,81)	30 (19,02)	22 (34,53)	22 (25,18)	27 (25,29)	**20 (25,71)**
Rest (N=72)	90 (22,33)	82 (25,69)	58 (29,16)	68 (33,90)	65 (30,25)	**76 (33,77)**

It is clearly shown that the blogs of Cluster 3, identified via the Cluster Analysis method of step 1, not only have the highest number of incoming links, but also the best ranking positions with a low standard deviation regarding all their Influence Indexes. Not only they are in most blogrolls of the rest of the blogs, but also they have the highest levels of readability and participation of the visitors which unavoidably leads to exerting higher influence on the community. These are the A-listed blogs stated earlier in the chapter. On the other hand, the ranking calculations showed that there are blogs among the 72 of Cluster 4 that possess very high rankings according to side influence indexes, as well as the overall Blog Authority Index. In addition to that, blogs with many incoming links of Cluster 1 and 2 appear to have low readability and influence, regardless the fact that the bloggers community redirects traffic towards them and therefore considers them as a homogenous group. For that reason, the third step is introduced as an attempt to further investigate the properties of the blogs of clusters 1, 2 and 4.

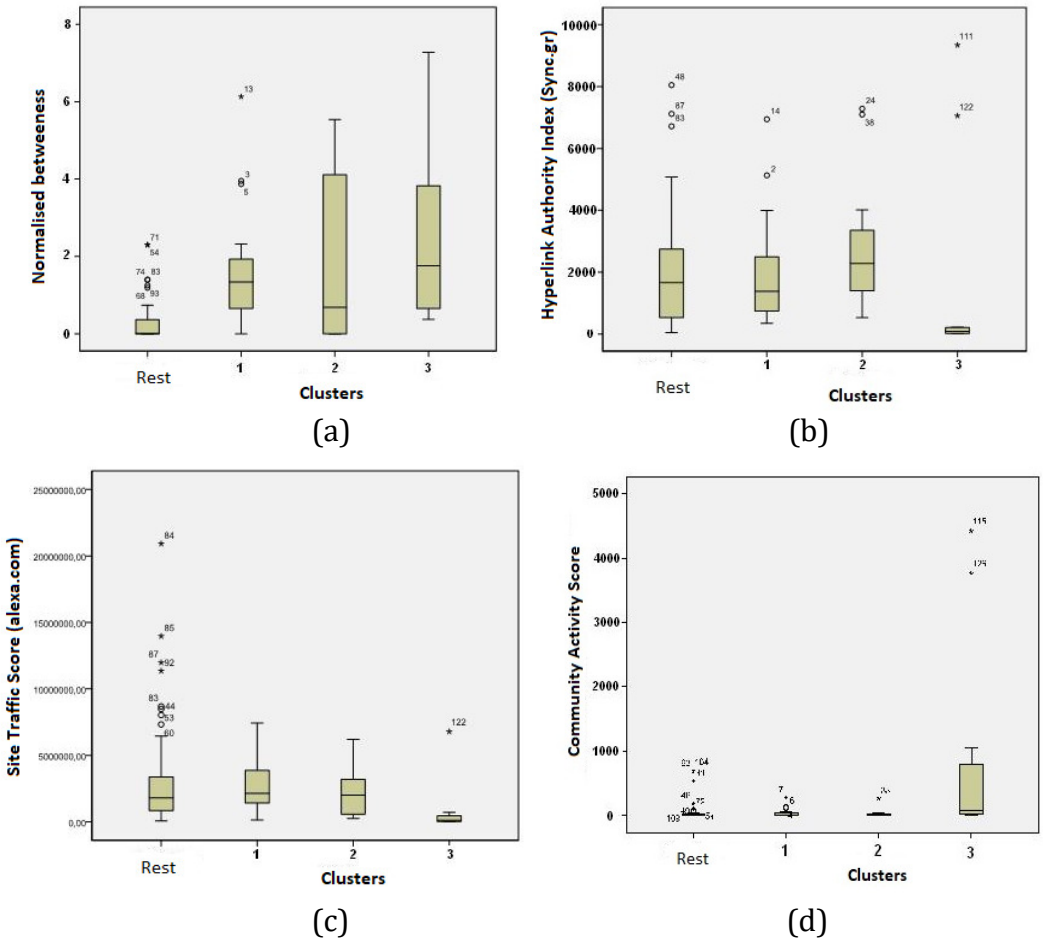

Figure 7: Boxplots of (a) Normalised betweeness, (b) Hyperlink Authority Index, (c) Site Traffic Score and (d) Community Activity Score for all Clusters (Zafiropoulos et al. 2012)

Figure 7 presents boxplots of data concerning all four indexes according to the 3 clusters and the rest 72 blogs found after applying the 1st component of the proposed model.

Component 3: Content Analysis

Although there has been a qualitative approach in the previous 2 steps of the analysis aiming at characterizing the blogs according to their content, however, a scientific approach is required to complement the method. For this purpose, Content Analysis (McMillan, 2000 andRichards, 2005) over a sample of the textual data of the blogs (List, 2003), has been applied. A 4500 words sample has been used, taken from clusters' 1,2 and 3 posts using a systematic sampling while the same sample has been taken from posts of the 7 most popular blogs of the remaining 72 according to their Blog Authority Index. The method has been partially applied in the previous steps when the comments have been counted as well as the hyperlinks. However, at this point it is about time to investigate thoroughly what the topics of discussion are (Hancock, 1998) and how they are related to the findings so far.

In this case study, QSR NVivo (version 8) software has been used for the Content Analysis. At first, a precise selection of nodes

(categories) has been made, which should be mutually excluding (Bazeley, 2009) and an overall coding over these nodes has been applied (kollias, 2006). The procedure has been done on three stages (McMillan, 2000), each one with a new set of nodes and each own coding. The qualitative properties of the blogs' content have been quantified by NVivo's queries which gave the final results. In order to produce reliable results, 2 independent coding procedures have been carried out and the corresponding results have been checked using the Kappa coefficient (Landis & Koch, 1977).

Prior coding the selection of nodes was decided for stage 1 (Herring, 2009) which was based on thematics deriving from e-participation literature. In this paper's approach they were taken by DEMO-net (www.demo-net.org) and they are the following 12: Campaigning, Community building/Collaborative Environments, Consultation, Deliberation, Criticism, Discourse, Electioneering, Information Provision, Mediation, Polling, Concern Creation, References, and Environmental Issues. Coding objectivity was validated by kappa coefficients with values greater than 0,718702 between the 2 independent coders for all the nodes of the first stage which corresponds to significant levels of agreement between (Landis & Koch, 1977). The blogs' data formed another group of nodes in NVivo (cases) which were assigned attributes like the cluster number they belong to (taken from step 1), the four side influence indexes along with their overall Blog Authority Index, the number of incoming hyperlinks and some objective parameters like the political orientation, the blog's scope and the writing style of its creator.

Regarding the second and third coding stage, the procedure was based on Emergent Coding. More specifically, the nodes selected for coding during the second stage were positive, neutral or negative references to all political parties and their leader of the period under investigation. Coding objectivity was also validated at this stage with kappa coefficients' values greater than 0,673002 for all 33 nodes of this stage.

The coding procedure of the third stage was based on a set of 32 nodes deriving by thematic areas of posts of all datasets of the 62 blogs whose content has been analyzed. The nodes referred to diverse issues prevailing in the period under investigation like Greece's debt and probable bankruptcy, criticism on the European Union and Euro, references to US, foreign Policy, nationalism, corruption or complication, memorandum and EU partners' negative attitude towards Greece as well as general news. After the coding of this stage the Kappa coefficients were again satisfactory for all nodes with a minimum value 0,775164.

The coding procedure of the three stages mentioned above led initially to quantitative results which referred to the existence or non-existence of a node and to the numbers of references of a node in a blog. Therefore, the qualitative properties of the blogs under investigation have been objectively underlined giving a better overview of the special characteristics of the A-list blogs or other interesting cases such as blogs with many incoming links and low influence or vice versa. Content Analysis results have also been used to qualitatively discriminate the blogs' clusters themselves. Therefore, the xi-squared test has been applied to determine whether there is a statistical difference of the existence of a node across the clusters ($p<0,05$). Statistically different values of the references of a node across the clusters ($p<0,05$) have been checked via ANOVA tests of mean values.

Taking a further step, it was statistically feasible at this point to use the 3-stages Content Analysis results to test the statistical significance of the clustering results of step 1. Principal Component Analysis has been used to reduce the dimensions of the initial observations and consequently, Discriminant Analysis has been applied for testing the

statistical significance of their classification into clusters of step 1 using Wilk's Lambda. This test has been carried out for both the nodes existence results in the blogs as well as for the total references of the nodes in the blogs. Discriminant Analysis of the first-stage coding showed that 59,7% (Wilks' Lambda=0,379) of the blogs have been identically classified into the clusters of step 1, taking into account the existence of nodes in blogs. 64,5% (Wilks' Lambda=0,216) and 61,3% (Wilks' Lambda=0,369) of blogs have been identically classified into the clusters of step 1 regarding second-stage and third-stage coding respectively. The results were even better when the number of nodes' reference was used for the Discriminant Analysis test. 74,793% (Wilks' Lambda=0,330), 66,1% (Wilks' Lambda=0,277) and 66,1% (Wilks' Lambda=0,288)- were the rates of the identical classification for the three stages of coding respectively (Table 7).

Table 7: Discriminant Analysis results of testing the statistical significance of clustering due to incoming hyperlinks and Content Analysis results by using Wilk's Lambda.

	Nodes in Blogs		Number of references of nodes in blogs	
	Clustering identification (%)	**Wilks' Lambda**	**Clustering identification (%)**	**Wilks' Lambda**
1st stage coding	59,7	0,379	74,793	0,330
2nd stage coding	64,5	0,216	66,1	0,277
3rd stage coding	61,3	0,369	66,1	0,288

Finally, the correlation of references' number of nodes across the blogs per cluster with the Influence measurements has been investigated by using Pearson Coefficients. The overall results are discussed in chapter 5.

Discussion

The multidimensional model of analyzing the political blogosphere introduced in this paper exploited the available big data from diverse perspectives. Component 1 attained to recognize three clusters plus one according to the number of incoming links and thus to the way the bloggers community envisages the blogosphere. It identified the A-list blogs which has been confirmed with Influence Analysis applied in Component 2, while it identified groups of blogs which were considered to be homogenous according to both; the incoming links and the analysis of their content applied in Component 3. Influence Analysis also demonstrated that incoming hyperlinks don't uniquely determine the influence a blog can exert highly influential blogs from the Internet community perspective were found, while they are not considered "central" in the bloggers' community. Cluster Analysis also succeeded in recognizing a whole group of blogs whose creators had similar political orientation (Cluster 1), which was scientifically confirmed by Content Analysis of step 3. Content Analysis also found similar characteristics (nodes) in blogs of Cluster 2,3 and the 7 highly influential blogs of the remaining 72. For example, Blogs of Cluster 2 included posts promoting mobilization and collective action and densely raising concerns related to political issues. Blogs' posts of Cluster 3 were the most objective regarding their criticism towards all political parties and leaders while they contained the

maximum characteristics (nodes) of the third stage coding. Discriminant Analysis applied in Step 3 certified that in the content of Cluster 1, Information Provision and Criticism were highly correlated with their betweenness in the blogosphere. In Cluster 2, on the other hand, Information Provision is negatively correlated with the blogs' betweenness. By analyzing the blogs' content and forming Cluster 3, it was found that Collaborative Community building was negatively correlated with their Hyperlink Authority Score which indicates the journalistic nature of these blogs that are used by visitors as alternative news sources and not as platforms for discursive collaboration.

From what stated above, it is clear that each component of the proposed method complemented and confirmed the findings of other components, a fact that underlines the necessity for a multifaceted approach for analysis (Gregory et al. 2011 and Shah & Sukthankar, 2011) of such a complicated and rich Big Data field like the political blogosphere, towards objective findings.

Conclusions and Further Work

Considering the subjectivity and the dynamic nature of the data used in this case study, the results of the above case study analysis recommend that the proposed method for analyzing the Activated Public Opinion in the Political blogosphere could be fertile terrain for further research and a potential method's integration over more advanced social media applications in Fields of Politics among other fields. The automatically implemented analysis (Cluster Analysis) seemed to give a satisfactory top level overview of the blogs under investigation but for a more detailed investigation, a more systematic approach entailing time consuming work is unavoidably required (Krishnamurthy, 2002). During the implementation several decisions and assumptions had to be made, all scientifically determined. Sampling had to be unavoidably done as this reduced the workload at feasible levels, but not at the expense of the accuracy of the final results.

The promising results of the current case study can encourage further research in this area adopting the proposed method over longer time windows with larger sets of Big Data. Future projects might focus on specific percussive political events and the time period before and after them. The blogs or other social networking nodes belonging to clusters that can automatically be found by applying Cluster Analysis (Park et al., 2005) might further be investigated by developing an Artificial Intelligence Software that could automatically produce Sentiment Analysis or any other automatic but accurate qualitative classification. Regarding the blogs themselves, these procedures can lead to the development of prediction models (Nallapati & Cohen, 2008) of the whole dynamically evolving blogs' social network, inevitably taking into consideration their interconnection with contemporary modern Social Networking sites.

References

- Adamic, L. & Glance, N. (2005), The political blogosphere and the 2004 U.S. election: divided they blog. In Proceedings of the 3rd international workshop on link discovery:36–43.
- http://www.blogpulse.com/papers/2005/AdamicGlanceBlogWWW.pdf
- Albrecht, S. Lübcke, M. & Hartig-Perschke , R. (2007), Weblog Campaigning in the German Bundestag Election 2005. Social Science Computer Review 25 (4): 504-520.
- Bazeley P. (2009), Analysing Qualitative Data: More than "Identifying Themes", Malaysian Journal of Qualitative Research, 2009.
- Borgatti, S.P., Everett, M.G. and Freeman, L.C. (2002), UCINET for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.
- Brin S., Page L., (2008), The Anatomy of a Large-Scale Hypertextual Web Search Engine, Computer Science Department,

Stanford University, Stanford, http://infolab.stanford.edu/~backrub/google.html
- Burnham, K. P. & Anderson, D. R. (2004), "Multimodel inference: understanding AIC and BIC in Model Selection", Sociological Methods and Research 33: 261–304.
- Chadwick, A. (2006), "Internet Politics: States, Citizens and New Communication Technologies", Oxford University Press, 2006.
- Du, H. & Wagner, C. (2006), Weblog success: Exploring the role of technology. International. Journal of Human-Computer Studies 64: 789–798
- Drezner, D. & Farrell, H. (2004), The power and politics of blogs, paper presented at the Annual Meeting of the American Political Science Association, Washington, DC, August. http://www.utsc.utoronto.ca/~farrell/blogpaperfinal.pdf
- Drezner, D. W., & Farrell, H. (2004), Web of influence. Foreign Policy, 145: 32–40.
- Drezner, D. & Farrell, H. (2008), The power and politics of blogs Public Choice 2008, 134:15–30.
- Efimova, L. & Hendrick, S. (2005), In search for a virtual settlement: an exploration of Weblog community boundaries. https://doc.novay.nl/dsweb/Get/Document-46041
- Graf, J. (2006), The Audience for Political Blogs. New research on Blog Readership. GW's Institute for Politics, Democracy & the Internet http://www.ipdi.org/uploadedfiles/audience%20for%20political%20blogs.pdf
- Gregory M., Engel D., Bell E., Piatt A., Dowson S. & Cowell A. (2011), Automatically Identifying Groups Based on Content and Collective Behavioral Patterns of Group Members, Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2787
- Hancock, B. (1998), Trent Focus for Research and Development in Primary Health Care: An Introduction to Qualitative Research. Trent Focus, University of Nottingham,
- Hanneman R.A., Riddle M., (2005), "Introduction to social network methods", Riverside, CA: University of California, published in digital form at http://faculty.ucr.edu/~hanneman/
- Herring, S. C., Kouper, I., Scheidt, L. A., & Wright, E. (2004), Women and children last: The discursive construction of weblogs. In L. Gurak et al. (Eds.), Into the Blogosphere: Rhetoric, Community an Culture of Weblogs,
- http://blog.lib.umn.edu/blogosphere/women_and_children.html
- Herring, S. C., (2009), "Web Content Analysis: Expanding the Paradigm", published in Herring, S. C., Kouper, I., Paolillo, J., Scheidt, L. A., Tyworth, M., Welsch, P., Wright, E., & Yu, N. (2005), Conversations in the blogosphere: An analysis "from the bottom up." Proceedings of the Thirty-Eighth Hawai'i International Conference on System Sciences. Los Alamitos: IEEE.
- Jackson, N. (2006), Dipping their big toe into the blogoshpere. The use of weblogs by the political parties in the 2005 general election. Aslib Proceedings: New Information Perspectives 58(4): 292-303
- Karpf, D. (2008), "Measuring Influence in the Political Blogosphere. Who is Winning and How Can We Tell?", George Washington University's Institute for Politics, Democracy and the Internet, 2008. http://www.the4dgroup.com/BAI/articles /PoliTechArticle.pdf.
- Keren, M. (2006) Blogosphere: the New Political Arena. New York: Lexington Books.
- Krishnamurthy, S. (2002), The Multidimensionality of Blog Conversations: The Virtual Enactment of September 11. In Maastricht, The Netherlands: Internet Research 3.0
- Landis, J.R. & Koch, G.G. (1977), "The measurement of observer agreement for categorical data". Biometrics, 33,1977: 159-174.

- McMillan, S. J. (2000), The microscope and the moving target: The challenge of applying content analysis to the World Wide Web. Journalism and Mass Communication Quarterly, 77(1): 80-98.
- Nallapati R. & Cohen W. (2008), Link-PLSA-LDA: A new unsupervised model for topics and influence of blogs. In Proceedings of the 2nd International Conference on Weblogs and Social Media.
- Park H., Thelwall M. & Kluver P. (2005), "Political Hyperlinking in South Korea: Technical Indicators of Ideology and Content", Sociological Research Online, Volume 10, Issue 3, http://www.socresonline.org.uk/10/3/park.html
- Reynolds, G. (2006), An Army of Davids: how markets and technology empower ordinary people to beat bigmedia, big government and other Goliaths. New York: Nelson.
- Richards, L. (2005) "Up and Running in your project: a post workshop for NVivo 7" published in "Handling Qualitative Data", Sage, London 2005
- Richards, L. (2005), Handling qualitative data. London: Sage.
- Shah F. & Sukthankar G. (2011), Using Network Structure to Identify Groups in Virtual Worlds, Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media,
- http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2805,
- Sigala, M. (2008). Web 2.0 tools empowering consumer participation in New Product Development: findings and implications in the tourism industry. Annual International International Council for Hotel, Restaurant and Institutional Education, (I-CHRIE) Convention "Welcoming a new era to hospitality education". Atlanta, Georgia, USA: 30 July – 2 August, 2008
- Trammell, K. & Keshelashvili (2005), A. Examining the new influencers: A self-presentation study of A-list Blogs. Journalism and Mass-Communication Quarterly 2005, 82, 4 :968-982.
- Vagianos, D., Goede, M. & Luca, V., (2019), Social Media Assisted Blog Content Dissemination: A two Case Studies Applied Analysis of Ruling Factors", accepted for publication in the Cyprus Journal of Sciences, Vol. 17, 2019.
- Wallsten, K. (2005), Political Blogs and the Bloggers Who Blog Them: Is the Political Blogosphere and Echo Chamber? Paper Presented at the American Political Science Association Annual Meeting Washington, D.C. September 1-4, 2005. http://www.journalism.wisc. edu/ blog-club/Site/Wallsten.pdf
- Wallsten, K. 2007. "Political Blogs: Transmission Belts, Soapboxes, Mobilizers or Conversation Starters?" Doctoral dissertation chapter, University of California, Berkeley
- Zafiropoulos, K. & Vrana, V. (2009), Representation and study of political blogs conversational patterns. 4th Mediterranean Conference on Information Systems Athens, Greece, 25-27 September 2009.
- Zafiropoulos, K. & Vrana, V. (2011), Hyperlink Analysis of Political Blogs Communication Patterns. USA: NOVA Publishers, 2011.
- Zafiropoulos, K., Vrana, V. and Vagianos, D. (2012), Bloggers' Community Characteristics and Influence within Greek Political Blogosphere. Future Internet 2012, 4, 396-412. https://doi.org/10.3390/fi4020396.