International Journal of Humanities and Management Sciences (IJHMS) Volume 1, Issue 2 (2013) ISSN 2320–4044 (Online)

# E-Government Exploration using Social Network Analysis Methods

Arbana Kadriu[1], Marina Saveska[2], Lejla Abazi-Bexheti[3]

***Abstract*—**The research presented in this paper uses the best methods for the analysis of social and information networks in order to achieve better and more efficient analysis of e-government. While there is growing interest in social networks, little attention has been paid to the application of this methodology in the field of e-government. Particulary, we analyze the web presence of the government in Macedonia and the relation of different government etnities with each other. Morever, we try to explore the relation between the government area with the non-government district. First, we create a database of interactions between web different etnities. The gained information is stored and represented by a graph. And then we explore this created network aiming to understand how electronic government interacts internally and externally. At the end we make a presentation of the results obtained from the analysis of the network where we can conclude which entity is the most popular, powerful, closest, important and central.

***Keywords*—**SNA metrics, e-government, centrality, PageRank

## I. Introduction

SOCIAL network analysis social network analysis focuses on the relationships and ties between actors within a network [1]. The last decade has witnessed a new movement in the study of social networks, with the main focus moving from the analysis of small networks to those with thousands or millions vertices, and with a renewed attention to the topology and dynamics of networks [2]. There are two reasons for that: the massively use and expansion of social media networks and availability of massive datasets that can be gained from these media.

There have been a few attempts of using differnet metrics to measure the role of the electronic government towards citizens.

Escher et.al. use methods from computer science (particularly webmetrics) and political science (a ‘tools of government’ approach) to go further than previous work in developing a methodology to quantitatively analyse the structure of government on the web [3].

Arbana Kadriu[1], Marina Saveska[2], and Lejla Abazi-Bexheti[3] are with CST Department, SEE-University 1200 Tetovo, Macedonia E-mail: a.kadriu@seeu.edu.mk[1], m.saveska@seeu.edu.mk[2], l.abazi@seeu.edu.mk[3]

In his paper, Marin aim is to analyze across European countries the impact of e-Government policies on their adoption, under different levels of Internet penetration, enabling an assessment of how promotion of e-Government (through investment in more and better services, for example) can have the maximum impact on citizenship adoption [4]. It reports analysis of a cross-sectional dataset of European countries using a Bayesian linear model.

Petricek & Escher describe preliminary work that examines whether statistical features of the structure of websites can be an informative measure of their quality [5].

Li et.al. in order to improve government transparency and promote interagency coordination, have investigated interagency networks. They show that World Wide Web (WWW) and Linked Open Data (LOD) can be used to automatically generate interagency networks and thus promote interagency network analysis to a larger scale and more practical status [6].

Adamic&Glance studied the linking patterns and discussion topics of political bloggers. Their aim is to measure the degree of interaction between liberal and conservative blogs, and to uncover any differences in the structure of the two communities [7].

The objective of this paper is to perform efficient research on e-government hyperlinks using methods for social network analysis. These methods help us to achieve efficient analysis of e-government in the Republic of Macedonia. This analysis is conducted using the tools NodeXL and VOSON.

## II. Social Network Analysis and Centrality Measures

In this research, we apply four standard centrality measures: degree centrality, closeness centrality, betweenness centrality, and PageRank [8].

For a given graph (network) G with N nodes and E edges, the main centrality measures are defined as follows:

### A. *Degree centrality.*

Degree centrality of a node is the number of unique edges that are connected to it. The equation of it is as following where $d(n)_i$ is the degree of $n_i$:

$$C_B(n_i) = \sum_{j,k \neq i} \frac{g_{ijk}}{g_{jk}}$$

where $g_{ijk}$ is all geodesics linking node *j* and node *k* which pass through node $i$ and $g_{jk}$ is the shortest path between the nodes j and k.

### B. *Closeness centrality*

Closeness centrality is a measure of the average shortest distance from each node of the network to each other node. If we denote the distance between node i and j with as $d(n_i,n_j)$, the closeness equation will be:

$$C_C(n_i) = \sum_i \frac{1}{d(n_i, n_j)}$$

### C. *Betweeness centrality*

Betweeness centrality is a measure of a node's centrality in a network equal to the number of shortest paths from all vertices to all others that pass through that node. Authors with high betweenness are the brokers and connectors who bring others together. Being between means that a vertex has the ability to control the flow of knowledge between most others. The betweenness centrality of a node v is given by the expression:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{Q_{st}(v)}{Q_{st}}$$

### D. *PageRank*

PageRank of a page X is the limiting probability that a random walk across hyperlinks will end up at X, as we run the walk for larger and larger numbers of steps [8]. This idea is summarized in the PageRank equation as follows:

$$PR(p) = (1-r)\frac{1}{N} + r\sum_{i=1}^{k}\frac{PR(p_i)}{C(p_i)}$$

where $N$ is the total number of pages, $r$ is a reducing factor, $C(p)$ is the outdegree of page $p$, and $p_i$ denotes the inlinks of $p$.

We define also the network metrics:

- **Distance** between a pair of nodes – is the edge length of the shortest path between them
- **Graph's diameter** - is the maximum of the shortest distances between node pairs
- **Graph's density** – measures how many edges are in graph compared to the maximum possible number of edges between vertices in that graph.

## III. CREATING NETWORK OF HYPERLINKS

We can view the static Web consisting of static HTML pages together with the hyperlinks between them as a directed graph in which each web page is a node and each hyperlink a directed edge. We refer to the set of all such nodes and directed edges as the web graph. Usually there is some text surrounding the origin of the hyperlink on page A. This text is generally encapsulated in the href attribute of the <a> (for anchor) tag that encodes the hyperlink in the HTML code of page A, and is referred to as anchor text . As one might suspect, this directed graph is not strongly connected: there are pairs of pages such that one cannot proceed from one page of the pair to the other by following hyperlinks. We refer to the hyperlinks into a page as in-links and those out of a page as out-links [10].

We have used VOSON system[1] to crawl and create the network of government hyperlinks. It is web-based software for the collection and analysis of hyperlink networks. Fig. 1 shows the initialization of crawling process, where we use seed list of initial hyperlinks. The gained network has 942 nodes. It includes hyperlinks of different government sites, as well as non-government ones, as are for example educational sites. Despite Macedonian sites, it contains sites from other countries worldwide, which do interact with the Macedonian government.

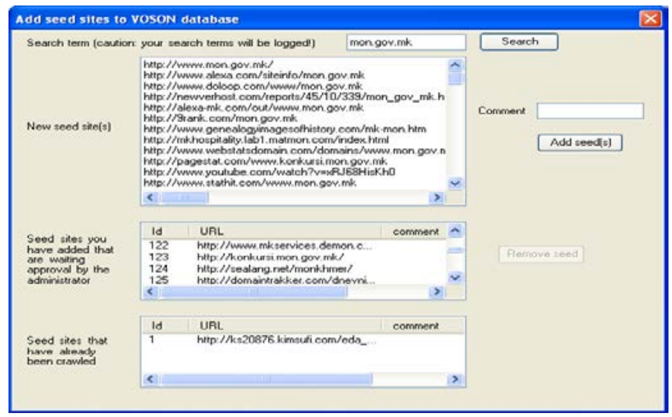


Fig. 1. Initialization of crawling process using VOSON

After the network creation, we exported it do the NodeXL software[2], which is a free, open-source system that makes it easy to explore network graphs.

## IV. DATA ANALYSIS AND RESULTS

### A. *Centrality Metrics*

The first element that we wanted to explore was the degree centrality, and it means to find which entity is the most "popular" in the network. We have gained that *Secretary for European Issues* (http://sep.gov.mk) and *State Labour Inspectorate* (http://www.dit.gov.mk) have higher popularity comparing to the other nodes. The first site has 185 and the second one has 135 "friends" in the obtained network.

Fig. 2 shows the visual representation regarding this metric – the most active nodes are colored as blue and less active are colored as pink. The above mentioned sites are annotated as blue squares.

The other aspect that we wanted to explore was which sites are the most powerful, which depends from their betweenness centrality values. Fig. 3. gives a description of the obtained outcome, where the elements with the highest betweenness centrality (8.00) are in orange, and those with lower value are in blue (2.00).

[1]You can find this software at: http://www.uberlink.com/

[2] http://nodexl.codeplex.com/

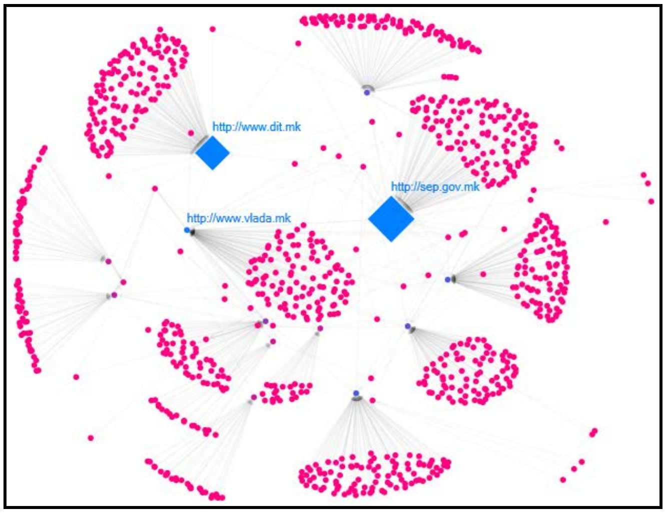


Fig. 2. Visual presentation for degree centrality

From this graph, it can be concluded that the most "powerful" entity is the site *Government of Republic of Macedonia* (http://www.vlada.mk). It has the highest influence in the network. The second in the list is *State Labor Inspectorate* (http://www.dit.gov.mk) and *Secretary for European Issues* (http://sep.gov.mk) is the third most powerful node.

The next investigated metric was about closeness centrality – we investigate the average shortest distance from each vertex to each other vertex. Actually, if a relevant government information needs to flow through the network, some entities would be able to get a message to all the other sites relatively quickly (i.e., in few steps), whereas others may require many steps. So, we explore which sites are nearer to the relevant information.

Fig. 4. gives an overview of the obtained metrics.

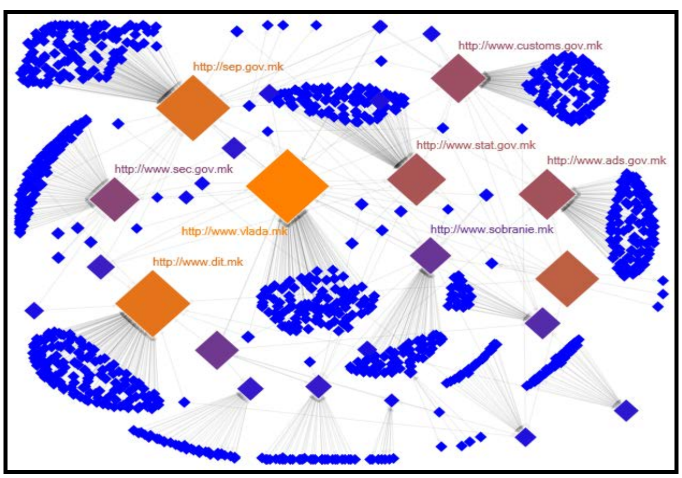


Fig. 3. Visual presentation for betweenness centrality

From this figure, it can be seen that the closest node in our network is the site *Government of Republic of Macedonia* (http://www.vlada.mk). Follow *Secretary for European Issues* (http://sep.gov.mk), *State Labor Inspectorate* (http://www.dit.gov.mk) and the site *Parliament of Republic of Macedonia* (http://www.sobranie.mk).

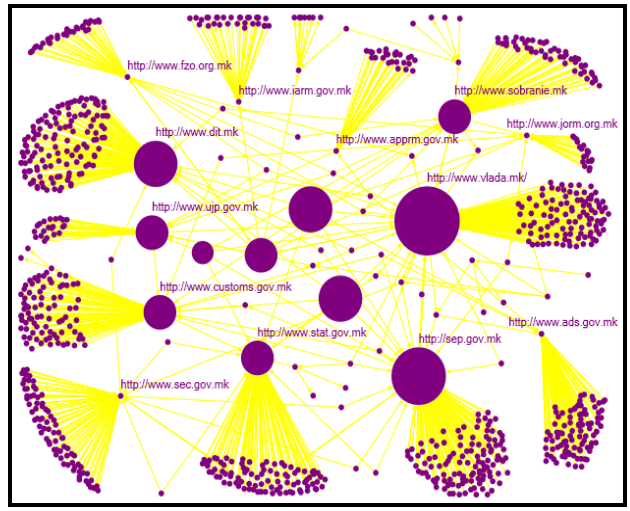


Fig. 4. Visual presentation for closeness centrality

Sites with low closeness centrality come out to be *Agency for Administration* (http://www.ads.gov.mk), *Securities Commission* (http://www.sec.gov.mk), *Agency for Entrepreneurship Support* (http://www.apprm.gov.mk), *Health Insurance Fund* (http://www.fzo.org.mk), *Institute for Accreditation* (http://www.iarm.gov.mk).

At the end, we calculate PageRank value for each hyperlink. With this metric, we get the most important nodes in our group of linked sites. In this way, the most important site in electronic government of Macedonia comes to be *State Labor Inspectorate*. The site with the lowest PageRank value is Macedonian Information Agency (http://www.mia.com.mk).

Table I gives an overview of different metrics for top 20 sites of our network.

### *B. Exploring the nature of network*

For our network we gained that:

- It's graph's diameter diameter is 7.
- Average distance is 3.93
- The graph density is 0,00113

The first two metrics means that our network satisfies the "small world"[3] phenomena [9]. The last metric indicates that the network has only a few edges – it is a sparse graph.

We have investigated the type of the hyperlinks that are linked with the government network. From them, 495 are of type *.com*, 109 of type *.org.*, 76 were *.gov*, 55 were *.net* and 25 were of type *.edu*.

The second interesting information is which countries are likely to be linked with the government of Macedonia. Table II gives the top countries. It is interesting that in this top list do not show Macedonia neighbors. Instead, there are countries form other continents such as USA (52), Australia (20), South

[3] "Small world" phenomena comprised few experiments conducted by Milgram and other researchers examining the average path length for social networks. The research was groundbreaking in that it suggested that human society is a small-world-type network characterized by short path-lengths.

Africa (5). This register can be analyzed in political context, finding out which countries are politically the most important for our country.

Most of the sites are naturally from Macedonia (77).

TABLE I
. METRICS FOR THE TOP 20 SITES

| Site | Deg | Betw. | Clos. | PageR |
|---|---|---|---|---|
| www.vlada.mk | 128 | 1,000 | 2,510 | 51,464 |
| sep.gov.mk | 138 | 0,874 | 2,684 | 58,258 |
| www.dit.mk | 135 | 0,887 | 2,721 | 59,958 |
| www.usaid.org.mk | 1 | 0,000 | 4,827 | 0,571 |
| www.usaid.org.mk | 1 | 0,000 | 4,827 | 0,571 |
| www.sobranie.mk | 62 | 0,416 | 2,830 | 25,574 |
| www.ads.gov.mk | 99 | 0,632 | 3,132 | 44,553 |
| www.apprm.gov.mk | 28 | 0,182 | 3,479 | 12,685 |
| www.fzo.org.mk | 37 | 0,207 | 3,755 | 15,201 |
| www.sec.gov.mk | 87 | 0,532 | 3,452 | 38,402 |
| status.katastar.gov.mk | 1 | 0.000 | 5,073 | 0,558 |
| www.iarm.gov.mk | 31 | 0,205 | 3,765 | 14,295 |
| www.stat.gov.mk | 97 | 0,662 | 2,880 | 42,617 |
| www.katastar.gov.mk | 4 | 0,021 | 4,074 | 1,921 |
| www.uslugi.gov.mk | 4 | 0,013 | 3,170 | 1,528 |
| e-administracija.gov.mk | 4 | 0,001 | 3,325 | 1,504 |
| prijava.aa.mk | 1 | 0,000 | 4,131 | 0,533 |
| www.president.gov.mk | 2 | 0,000 | 3.424 | 0,856 |
| www.mia.com.mk | 1 | 0,000 | 3,509 | 0,500 |
| www.mio.gov.mk | 2 | 0,000 | 3,508 | 0,819 |
| www.jorm.org.mk | 24 | 0,140 | 3,091 | 9,960 |

Table II
List Of Top States Linked With The Government

| State | Number of linked states |
|---|---|
| **Macedonia** | 73 |
| **USA** | 52 |
| **United Kingdom** | 30 |
| **Denmark** | 27 |
| **Australia** | 20 |
| **Holland** | 13 |
| **Czech Republic** | 10 |
| **Canada** | 9 |
| **Sweden** | 9 |
| **Russia** | 8 |
| **Belgium** | 7 |
| **Germany** | 7 |
| **Slovakia** | 7 |
| **South Africa** | **5** |
| **Ireland** | 4 |

## V. CONCLUSION

We described a methodology and results for analyzing the work of electronic government. This approach can be replicated for other groups, which can be non-government networks, business or educational groups etc.

## REFERENCES

[1] R. Breiger, "The Analysis of Social Networks," in Handbook of Data Analysis, edited by M. Hardy and A. Bryman. London: Sage Publications, 2004, pp. 505–526.

[2] M. Newman. (2001a), "Scientific collaboration networks: I. Network construction and fundamental results", Physical Review E, 64, 016131

[3] Tobias Escher, Helen Margetts, Vaclav Petricek, Ingemar Cox, "Governing from the Centre? Comparing the Nodality of digital Governments", Annual Meeting of the American Political Science Association, Philadelphia, 31 Aug - 4 Sept, 2006.

[4] Havier Fernandez, Marin, "The Impact of e-Government Promotion in Europe: Internet Dependence and Critical Mass", Institut Barcelona d'Estudis Internacionals (IBEI), Volume 3, Issue 4, 29-December 2011.

[5] Vaclav Petricek, Tobias Escher, Ingemar J.Cox, Helen Margetts, "The web structure of E-government-Developing a methodology for Quantitative Evalution", In Proceedings of the 15th International Conference on World Wide Web (Edinburgh, Scotland, May 23 - 26, 2006). WWW '06. ACM Press, New York, NY, 669-678.

[6] Ding, Li and Whalen, Ryan and McGuinness, Deborah L. and Contractor, Noshir (2011), "Government Agency Network Discovery via Hyperlinks and Datalinks", pp. 1-3. In: Proceedings of the ACM WebSci'11, June 14-17 2011, Koblenz, Germany.

[7] Lada A. Adamic and Natalie Glance (2005), "The Political Blogosphere and the 2004 U.S. Election: Divided They Blog", LinkKDD-2005, Chicago, IL, Aug 21, 2005.

[8] D. Easley and J. Kleinberg. Networks, Crowds, and Markets: Reasoning about a Highly Connected World. Cambridge University Press, 2010.

[9] D. Watts, S. Strogatz, "Collective dynamics of 'small-world' networks", Nature 393 (6684): 409–10, 1998.

[10] Christopher D. Manning. Prabhakar Raghavan, Hinrich Schütze (2008), *Introduction to Information Retrieval,* Publisher Cambridge University Press.