Visualizing the Evolution of Groups of Politicians Mentioned in the News

Carolina Bento*‡, Daniel Gonçalves†‡ Bruno Martins*‡
* *Data Mining and Information Retrieval Group, INESC-ID,*
† *Visualization and Intelligent Multimodal Interfaces Group, INESC-ID,*
‡ *Instituto Superior Técnico, Universidade de Lisboa*
{*carolina.bento, daniel.j.goncalves, bruno.g.martins*}*@tecnico.ulisboa.pt*

***Abstract*—In this paper, we describe a tool to visualize and explore communities of politicians arising from mentions in news articles, combining techniques from network analysis and visual analytics. With this tool one can explore the communities, their organization according to parties, with whom a politician was mentioned with in a given year, or key political events, like the emergence of a new prime-minister. User tests involving our tool, and a collection of Portuguese news articles for a period of 10 years, have shown the interest on this kind of tool, and its applicability to analyze group evolution.**

***Keywords*-Community Visualization; Visualizing Community Evolution; Network Analysis**

I. Introduction

Analyzing newswire streams is significantly important for studies related to political science, since the area focuses on the analysis of entities that are widely covered in the media.

We propose a tool to analyze how politicians are mentioned in news articles, and to see their organization in terms of groups that are frequently mentioned together, i.e., *communities*. By studying the corresponding *mentioning patterns*, the *communities* of politicians, and how they *evolve over time*, i.e., grow, merge, split, etc., one can assess, for instance, if politicians tend to be more frequently mentioned with a peers from their party, or from other parties, and also, study their *presence* and continuity in *communities* over time, thus exploring the impact of political events.

A recent work, closely related to ours, is that of Bandari et al., in which the authors developed a method to determine group behavior based on user content preferences [1]. Their methodology was applied to political discourse, on the social news site *balatarin.com*. This site mainly contains with discussions in the Persian-language, and it was pivotal during the Iranian post-election in 2009 to spread information, and to propose and discuss ideas. Bandari et al. detected the communities of users with similar voting patterns in each yearly time-frame, from 2006 to 2010, and captured the evolution of each community with a *transition probability matrix* that encoded the probability of a community, from one time-frame, becoming another community in the consecutive time-frame. A visualization for paths of opinion-based communities is based on the highest transition probabilities from consecutive time-frames.

Our visualization tool can be used by social scientists, as well as the general public, to explore politicians' *mentioning patterns* in the news. One can have an overview on the number of *communities* per year, as well as their dimensions and dominant party (i.e., the party of the majority of its members), and perform tasks such as inspect each *community* and view its members, or directly assess the *presence* of a politician in the 10 year period.

We identified the *communities* using the Clique Percolation Method (CPM), which defines communities as k-cliques, i.e., unions of all complete sub-graphs with size equal to k that share $k-1$ nodes [2]. Then, we used the group evolution discovery method (GED), to determine *community* evolution [3]. To express the evolution of groups through time, we used the approach from Bródka et al. [3], considering a set of events that can change the state of *communities* between consecutive time frames.

II. Dataset and Community Characterization

We used a corpus of news articles from the politics section of the Portuguese newspaper Público, ranging from 2001 to 2010. The politicians, mentioned in the articles, were extracted with a Named Entity Recognition system, to create a *dataset* with the co-occurrence of politicians that is capable of reflecting their *mentioning patterns*. The *dataset* was split into yearly periods, i.e., considering only the relationships within a year. Then, we created *co-occurrence graphs*, where nodes represent politicians, and edges represent pairs of nodes that were mentioned in the same news article, weighted with their co-occurrence *frequency*.

To obtain *communities* reflecting the most frequent (i.e., relevant) relationships between politicians we only considered those appearing at least 10 times in the entire dataset, at least 5 times in the considered year, and only the pairs

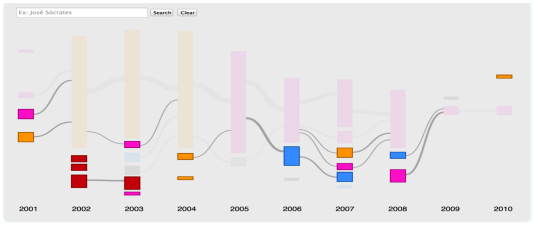

Figure 1: Overview on the communities from 2001 to 2010.

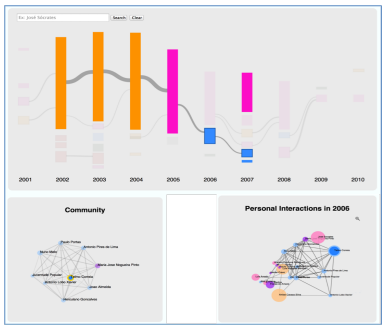

Figure 2: Community *presence* of Telmo Correia (top), and his community view and personal interactions (bottom).

of politicians that co-occurred at least 5 times in that year. We applied the CPM method to extract the 4-clique *communities*, and the GED method to analyze the evolution of each *community* in the 10-year period.

In each *community* all politicians were associated with their **party**, which allowed an analysis of *community* **homogeneity**. An homogeneous *community* comprises only politicians from the same party. Politicians were also characterized by their direct relationships, i.e., **personal interactions**, in the given year, and by their **social position**, i.e., their importance relative to the other *community* members, given by a weighted version of the PageRank algorithm [4]. We also attempted to identify the *community* leaders. Each *community* encodes the **strength** of its ties, given by edge weights. We also represent the *community*'s **dominant party**, i.e., the party with the majority of members in that *community*.

III. Visualizing Communities of Politicians

The proposed tool provides a main area with an overview on all the *communities* that were identified in the 10-year period – see Figure 1. Community evolution is represented as an interactive Sankey diagram, to emphasize the flow of politicians from *communities* in consecutive years. *Communities* are represented as bars, while the links connecting them encode one of the events in the GED method, i.e., the flow of politicians between them. The height of the bars is proportional to the number of members in the *community*, while the width of the links is proportional to the number of politicians that *migrated* from one community to the other. Some *communities* are also highlighted, while others are faded, indicating their homogeneity.

The *community* view uses an undirected graph encoding the members' **social position** (given by the size of the nodes), their relationships, and the strength of their ties (given by the thickness of the edges). The *community* leader stands out by having a gold ring and a darker color.

When viewing the *personal interactions* of a politician in a given year (Figure 2), the politician stands out, with an increased node size and a darker color. As we explore a politician in his *community*, all the *communities* which he was part of, and the *flow* between them, are highlighted, the same happening when we use the search box.

We conducted initial user tests with a group of 10 users (70% male and 30% female, with ages between $25-34$ (80%) and $55-64$ years old (20%). The users performed a set of tasks, namely (i) explore a given community, (ii) identify its members and its leader, and (iii) check the presence of a given politician. The last task allowed us to assess if the user understood the idea of community evolution. All users were able to perform all tasks, and they indeed acknowledged the usefulness of the tool. Several of them commented on interesting patterns discovered during the tests. However, they also pointed out some shortcomings. They mentioned that if they were to use this tool without indication of the tasks to perform, they would need more context. They referred that an overview text (e.g., FAQ), detailing the objectives of the tool, and the meaning behind communities, as well as a list of possible tasks, would support a better user experience, when without guidance.

IV. Conclusions

In this paper we describe a visual analytics tool that leverages on network analysis methods to explore communities extracted from co-occurrences in text, and how they evolve over time. We analyzed communities of politicians that are frequently mentioned together in news articles. However, this kind of tool can be used in other domains.

This tool can point out political events such as the emergence of politicians, e.g., *Pedro Passos Coelho*, who only became frequently mentioned in news articles in 2008, coinciding with his *rise* to become Portugal's prime-minister in 2011. We could also verify that the party in office in a given year is usually the dominant party of the majority of the communities from that year.

Interesting future work involves giving more meaning to the communities and to the relationships between politicians, e.g., by using tag clouds with the *topics* that tie communities together, and by contextualizing each year's political scene with the headlines of articles related to the communities.

Acknowledgments

This work was supported by Fundação para a Ciência e Tecnologia, through the project grants with references UTA-Est/MAI/0006/2009 (REACTION), EXPL/EEI-ESS/0427/2013 (KD-LBSN), and PEst-OE/EEI/LA0021/2013 (INESC-ID plurianual funding).

References

[1] R. Bandari and H. Rahmandad and V. P. Roychowdhury. Blind Men and the Elephant: Detecting Evolving Groups in Social News. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, 2013.

[2] G. Palla and I. Derényi and I. Farkas and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043), 2005.

[3] P. Bródka and S. Saganowski and P. Kazienko. GED: the method for group evolution discovery in social networks. In *Proceedings of the International Conference on Advances in Social Networks Analysis and Mining*, 2001.

[4] L. Page and S. Brin and R. Motwani and T. Winograd. The PageRank citation ranking: Bringing order to the Web. In *Proceedings of the International World Wide Web Conference*, 1998.