

## A Survey Paper on Automatic Extraction of Policy Network Using Web link and Documents.

**Mrs. Vaishali Chaudhari**

*G.S.Moze College of Engineering, Balewadi, Pune-45.  
University Of Pune, Pune, India.*

**Prof. J. Ratanaraj Kumar**

*G.S.Moze College of Engineering, Balewadi, pune-45.  
University Of Pune, Pune, India.*

### Abstract

*In Today's World Policy networks are used by economists and political scientists. They want to explain various financial and social phenomena, such as the development of partnerships between political entities or institutions from different levels of governance. The analysis of policy networks demands a series of arduous and time-consuming manual steps including interviews and questionnaires. We estimate the strength of relations between actors in policy networks using features extracted from data harvested from the web. Features include webpage counts, outlinks, and lexical information extracted from web documents or web snippets. In this paper we approach is automatic and does not require any external knowledge source, other than the specification of the word forms that correspond to the political actors. The features are evaluated both in isolation and jointly for both positive and negative actor relations. Correlation of up to 0.74 is achieved for positive relations.*

*Keywords— Policy Network, web link, web Documents, correlation, mean square error.*

### 1. Introduction

Policy networks are one well defined domain for research. Political scientists use policy networks to investigate social and financial phenomena, especially, the evolution of relations between actors and the effectiveness of policies toward the formation of partnerships among actors. This is achieved by

reference to the structure of networks in a given policy field at different phases of policy development. A policy network can be described by its actors, their linkages, and its boundary [2]. Policy networks consist of a set of public and private actors and a number of linkages between them that serve as channels for communication and the exchange of information, expertise, trust, and other policy resources. The network boundaries are not primarily determined by formal institutions but rather by functional relevance and structural embeddedness [2].

Policy networks are identified through a manual procedure performed by experts. Identifying actor's links, and boundaries, i.e., analysing a policy network's structure requires refined techniques and extensive and time-consuming manual collection of data through interviews and questionnaires. During the manual identification of networks, many subjective factors may be present, because this procedure relies strongly on the human subjects that participate in the interviews. Such factors include personal opinions, the person's willingness to participate, and even cultural issues.

Overall, policy network identification currently requires a "large scale investment" that does not always "lead to breathe taking empirical and theoretical results" [2]. When lacking the resources for data collection and network analysis, political scientists often revert to qualitative analysis or construct the network topology using their intuition, significantly limiting the evidence-based validation of their results. We have presented different methods to overcome issues related efficient automatic extraction of policy networks. Recently we have studied the method

presented in [1]. In this method author presented an algorithm for the automatic extraction of policy networks using information collected from the web. Here we are discussing more about this method presented in [1], because in this project we are later going to improve this method. In [1], the degree of relatedness (strength of link) between policy actors in a network is computed using three types of features on documents or snippets downloaded by web search engines, namely: 1) the frequency of co-occurrence for each pair of actors (in web documents), 2) the lexical contextual similarity between snippets of web documents in which the actors appear, and 3) the co-occurrence of hyperlinks present in web documents that contain the actors. For each type of feature and for their combinations, a variety of similarity metrics are used to estimate the link strength for each pair of actors. This presented approach in [1] was not intended to substitute expert knowledge, but rather it should be viewed as a low-cost, semi-automated computational tool that can significantly support and enhance policy network analysis. The proposed method aims to be efficient and reduce human biases.

## 2. Literature Survey

In the literature we have presented different methods and approaches those are used for political dataset analysis. More specifically, political analysts have used text mining to analyse electoral campaigns, identify voters' profiles, determine ideological positions, code political interaction, and detect political conflict's content [3], [4]. Textual data mostly consist of political manifestos, but transcribed speeches and political statements are also used. In [5], [6], the WORDSCORES system is proposed that extracts economic and social policy dimensions based on word frequencies from manifestos.

Similarly, the WORDFISH system [7] mines policy dimensions of parties and estimates their uncertainty over time using word frequencies from manifestos. Opinion mining is an active research area that is also relevant to political scientists. Opinions can be mined from text, blogs or from transcribed speech, e.g., [8]. Important research questions include the selection of lexical features (words and terms), the scores assigned to each term, as well as, the computational model used to combine the evidence, e.g., [9]. In [10], lexical features are combined with social information extracted from blog to classify political sentiments during the 2008 US Presidential election.

In [11], opinion mining techniques (including lexical feature selection) are applied to the analysis of political conflicts. Regarding social network analysis, political analysts have used network analysis to study formal and informal interactions. Policy networks extraction can be considered as a special type of social networks extraction, an active research area. The major steps in the extraction of social networks are relation identification [12], i.e., to identify whether two actors are related, relation labelling [16], [17], assign an existing relation to a category and the estimation of strength [18], i.e., identify whether an existing relation is weak or strong.

In [12], a network of experts with respect to certain topics is constructed by estimating similarity of users according to the frequency of co-occurrence of their names in web documents. Similarly, in [19] web co-occurrence of entities is used for creating a network of research communities. In [13] web co-occurrences are used for the extraction of social network of conference participants; a machine learning approach is used to classify each relation from a predefined set of relation types. In [16], automatically extracted key phrases are used to describe the relations between entities. E-mail contacts are used as features in [21] to create personal and professional relationship networks. In [23], social networks are extracted and updated over time using monolingual or multilingual news from articles. In [17], social networks of entities are extracted using posts from the blogosphere and the lexical context of entity pairs is used to automatically label the relations. In quoted phrases from novels are used to extract the social network of the novel's characters. In [1], author estimated the strength of relations between actors in policy networks using features extracted from data harvested from the web. Features include webpage counts, outlinks, and lexical information extracted from web documents or web snippets. The proposed method presented in [1] was automatic and does not require any external knowledge source, other than the specification of the word forms that correspond to the political actors. The features are evaluated both in isolation and jointly for both positive and negative (antagonistic) actor relations.

## 3. Existing System

The method which is presented recently for the automatic extraction of policy networks using the web links and documents is nothing but first step towards creating algorithms and tools useful to policy network analysts, therefore there few limitations associated with

this method [1]. The existing method presented in [1] was evaluated on two case studies with good results achieving correlation of up to 0.74 for positive relations. But it was shown that it is much harder to extract negative relations, and hence only moderate success was achieved for this task. Another problem associated with this method was among the metrics there was not a clear winner. A variety of parameters such as data sparseness, actor name ambiguity, Language, and relation type affect the performance of the relatedness metrics. In addition to this it is required to work over automatically identifying actors participating in policy networks and their lexicalizations.

### 3.1. Limitations of Existing Methods

Present method delivers good results for positive relations extraction whereas worst in case of negative relations extraction. This method and its performance metrics not properly evaluation in order to claim the efficiency of proposed approach. Several ambiguities were presented. Lack of automatic identification of actors.

## 4. Proposed Solution

In this approach we presenting improve method for automatic actor identification of actors those are participating in policy networks and their lexicalizations as well as Extraction of Policy Networks Using Web Links and Documents. The main focus of this work is to improve the performance for correlation in case both positive as well as negative relation extractions. In this project we are presenting the framework which will automatically compute the strength of relations between actors to automatically create policy networks. Different features extracted in this work that used information automatically extracted from the World Wide Web. Specifically, we investigated the use of page counts, lexical context, and outlinks, as well as, their fusion, as potential features for estimating relatedness between actor pairs. In addition to this efficient technique is presented for the automatically identifying actors participating in policy networks and their lexicalizations.

## 5. System Requirement & Specification

### 5.1. Software Requirements

Front End: Java  
Tools Used: Eclipse/Net beans  
Operating System: Windows 7

### 5.2. Hardware Requirements

Processor: Pentium IV 2.6 GHz  
Ram: 512 Mb  
Monitor: 15" Colour  
Hard Disk: 20 GB  
Floppy Drive: 1.44 Mb  
Keyboard: Standard 102 Keys  
Mouse: 3 Button

## 6. Conclusions

In conclusion, it is possible to automatically compute the strength of relations between actors to automatically create policy networks. A variety of features were proposed and evaluated that used information automatically extracted from the World Wide Web. Specifically, we investigated the use of page counts, lexical context, and outlinks, as well as, their fusion, as potential features for estimating relatedness between actor pairs. The proposed method was evaluated on two case studies with good results achieving correlation of up to 0.74 for positive relations. However, it was shown that it is much harder to extract negative relations, only moderate success was achieved for this task. Among the metrics there was not a clear winner. A variety of parameters such as data sparseness, actor name ambiguity, language, and relation type affect the performance of the relatedness metrics. The automatically extracted networks were also validated by political scientists and useful conclusions about the evolution of the networks over time were drawn.

## 7. REFERENCES

[1] Theodosia Moschopoulou, Elias Iosif, Student Member, IEEE, Leeda Demetropoulou, Alexandros Potamianos, Senior Member, IEEE, and Shrikanth (Shri) Narayanan, Fellow, IEEE, "Toward the Automatic Extraction of Policy Networks Using Web Links and Documents", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 10, OCTOBER 2013.

- [2] P. Kenis and V. Schneider, *Policy Networks and Policy Analysis: Scrutinizing a New Analytical Toolbox*, pp. 25-59, Westview Press, 1991.
- [3] L. Zhu, *Computational Political Science Literature Survey*, <http://www.personal.psu.edu/luz113>, 2013.
- [4] B. Monroe and P. Schrodt, "Introduction to the Special Issue: The Statistical Analysis of Political Text," *Political Analysis*, vol. 16, no. 4, pp. 351-355, 2008.
- [5] M. Laver, K. Benoit, and J. Garry, "Extracting Policy Positions from Political Texts Using Words as Data," *Am. Political Science Rev.*, vol. 97, no. 2, pp. 311-331, 2003.
- [6] K. Benoit and M. Laver, "Estimating Irish Party Policy Positions Using Computer Word scoring: The 2002 Elections - A Research Note," *Irish Political Studies*, vol. 18, no. 1, pp. 97-107, 2003.
- [7] J.B. Slapin and S.-O. Proksch, "A Scaling Model for Estimating Time-Series Party Positions from Texts," *Am. J. Political Science*, vol. 52, no. 3, pp. 705-722, 2008.
- [8] M. Thomas, B. Pang, and L. Lee, "Get out the Vote: Determining Support or Opposition from Congressional Floor-debate Transcripts," *Proc. Conf. Empirical Methods in Natural Language Processing*, pp. 327-335, 2006.
- [9] B. Chen, L. Zhu, D. Kifer, and D. Lee, "What Is an Opinion About? Exploring Political Standpoints Using Opinion Scoring Model," *Proc. 24th AAAI Conf. on Artificial Intelligence*, pp. 1007-1012, 2010.
- [10] W. Gryc and K. Moilanen, "Leveraging Textual Sentiment Analysis with Social Network Modelling: Sentiment Analysis of Political Blogs in the 2008 U.S Presidential Election," *Proc. 'From Text to Political Positions' Workshop*, 2010.
- [11] B. Monroe, M. Colaresi, and K. Quinn, "Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict," *Political Analysis*, vol. 16, no. 4, pp. 372-403, 2008.
- [12] H. Kautz, B. Selman, and M. Shah, "The Hidden Web," *AI Magazine*, vol. 18, no. 2, pp. 27-36, 1997.
- [13] Y. Matsuo, J. Mori, and M. Hamasaki, "POLYPHONET: An Advanced Social Network Extraction System from the Web," *Proc. 15th Int'l World Wide Web Conf.*, pp. 397-406, 2006.
- [14] H. Tomobe, Y. Matsuo, and K. Hasida, "Social Network Extraction of Conf. Participants," *Proc. 12th Int'l World Wide Web Conf.*, 2003.
- [15] Y. Jin, Y. Matsuo, and M. Ishizuka, "Extracting Social Networks among Various Entities on the Web," *Proc. European Conf. The Semantic Web: Research and Applications*, pp. 251-266, 2007.
- [16] J. Mori, T. Tsujishita, Y. Matsuo, and M. Ishizuka, "Extracting Relations in Social Networks from the Web Using Similarity Between Collective Contexts," *Proc. Fifth Int'l Semantic Web Conf.*, pp. 487-500, 2006.
- [17] F. Mesquita, Y. Merhav, and D. Barbosa, "Extracting Information Networks from the Blogosphere: State-of-the-Art and Challenges," *Proc. Fourth Int'l Conf. Weblogs and Social Media, Data Challenge Workshop*, 2010.
- [18] R. Xiang, J. Neville, and M. Rogati, "Modeling Relationship Strength in Online Social Networks," *Proc. 19th Int'l World Wide Web Conf.*, pp. 981-990, 2010.
- [19] P. Mika, "Flink: Semantic Web Technology for the Extraction and Analysis of Social Networks," *J. Web Semantics*, vol. 3, no. 2, pp. 211-223, 2005.
- [20] P. Mika, "Ontologies Are Us: A Unified Model of Social Networks and Semantics," *J. Web Semantics*, vol. 5, no. 1, pp. 5-15, 2007.
- [21] A. Culotta, R. Bekkerman, and A. McCallum, "Extracting Social Networks and Contact Information from Email and the Web," *Proc. First Conf. Email and Anti-Spam*, 2004.
- [22] A. Gruzd and C. Haythornthwaite, "Automated Discovery and Analysis of Social Networks from Threaded Discussions," *Proc. Int'l Network of Social Network Analysis*, 2008.
- [23] B. Poulliquen, R. Steinberg, and J. Belyaeva, "Multilingual Multi-Document Continuously-Updated Social Networks," *Proc. Int'l Conf. in Recent Advances in Natural Language Processing*, pp. 25-32, 2007.