ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

Automatic Extraction of Policy Networks from Social Networks and Snippet

Jayalakshmi.D[1], K.Sasikumar[2],

M.E, Department of Computer Science and Engineering, Mother Terasa College of Engineering and Technology, Mettusalai, Illuppur, Pudukkottai (dt), Tamil Nadu, India.

Head of the Department, Department of Computer Science and Engineering, Mother Terasa College of Engineering and Technology, Mettusalai, Illuppur, Pudukkottai (dt), Tamil Nadu, India.

ABSTRACT: Policy networks are widely used by political scientists and economists 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. In this system, 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. The proposed 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 (antagonistic) actor relations. The proposed algorithms are evaluated on two EU policy networks from the political science literature. Performance is measured in terms of correlation and mean square error between the human rated and the automatically extracted relations. Correlation of up to 0.74 is achieved for positive relations. The extracted networks are validated by political scientists and useful conclusions about the evolution of the networks over time are drawn.

KEYWORDS: Policy networks, social networks, relatedness metrics, similarity metrics, web search, policy actors, link analysis.

I. INTRODUCTION

Modern democratic governance reflects a shift away from the traditional notions of hierarchy toward more cooperative forms of public policy making. Within this context, the term "network" is often used to describe clusters of different types of actors, who are related in the political, social, and economic spheres. In earlier systems, the term "policy network" is defined as "a cluster of actors, each of which has an interest, or "stake" in a given policy sector and the capacity to help determine policy success or failure." 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 (planning, implementation, and evaluation). A policy network can be described by its actors, their linkages, and its boundary. 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. Typically, policy networks are identified through a manual procedure performed by experts. Identifying actors, links, and boundaries, i.e., analyzing 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 breathtaking empirical and theoretical

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

results". 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.

II. POLICY NETWORKS

A policy network can be considered as a type of social graph in which the nodes represent the actors involved in a given policy field, while their relations are represented by the edges. In social networks, nodes usually correspond to persons and edges represent relations among them built on a ground of mutual understanding, which can take several forms, such as friendship and coauthorship. In policy networks, actors can be organizations or even groups or unions of variable size and degree of formal organization. The relations among such actors usually signify the development of partnerships rather than a lax social relation. Furthermore, relations in policy networks depend on external factors, such as economic policies and funding at the local, national or supra-national level. Relations between policy actors (much like economic actors) can also be antagonistic rather than cooperative, or follow a more complex pattern of both cooperation and competition (sometimes referred to by economists as competition). Often policy networks are studied at their infancy when the links between actors are just emerging and might not be directly observable through common action or direct communication.

All these subtleties imply that established features and algorithms for social network analysis might not be directly applicable to policy network extraction. To our knowledge, this is the first comprehensive research effort toward the automatic extraction of policy networks. A variety of features extracted from web documents or snippets are proposed for estimating the relations between policy actors; these features are motivated by recent research in the fields of information retrieval and natural language processing. Another important contribution is that the proposed features and algorithms are evaluated against actual policy networks identified by expert political scientists. It is shown that the automatically extracted policy networks are capable of capturing the main relations between policy actors and are in broad agreement with networks built manually. In some cases, the automated method is shown to offer a deeper understanding of the relations between actors, especially as this pertains to the evolution of a network over time and policy outcomes.

III. EXISTING APPROACHES – A SUMMARY

The use of computational analysis of large amounts of data by political analysts has flourished in the past few decades facilitating the study of group connections. Two computational methods have been widely used in political science namely text analysis and social network analysis. More specifically, political analysts have used text mining to analyze electoral campaigns, identify voters' profiles, determine ideological positions, code political interaction, and detect political conflict's content.

Textual data mostly consist of political manifestos, but transcribed speeches and political statements are also used. In earlier approaches, the WORDSCORES system is proposed that extracts economic and social policy dimensions based on word frequencies from manifestos. Similarly, the WORDFISH system 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. 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.

In earlier systems, lexical features are combined with social information extracted from blog to classify political sentiments during the 2008 US Presidential election. In earlier systems, 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, i.e., to identify whether two actors are related, relation labeling, assign an existing relation to a category and the estimation of strength, i.e., identify whether an existing relation is weak or strong.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

The most common feature used to identify a relation is the frequency of co-occurrence of the related pair of terms in web documents, but other features, such as, lexical context, keyphrases, log files, and e-mail information are also used. In earlier systems, 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 earlier systems, web co-occurrence of entities is used for creating a network of research communities.

In earlier systems, 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 earlier systems, automatically extracted key phrases are used to describe the relations between entities. E-mail contacts are used as features in earlier systems to create personal and professional relationship networks. In earlier systems, social networks are extracted and updated over time using monolingual or multilingual news from articles. In earlier systems, 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 earlier systems, quoted phrases from novels are used to extract the social network of the novel's characters. In earlier systems, the log files of shared workspaces are used to extract user-oriented and objectoriented social networks. Some of the aforementioned systems also apply generic relatedness metrics used in other fields, such as Natural Language Processing and Information Retrieval. For example, metrics that are based on web co-occurrence and lexical features are extensively used for the computation of semantic similarity between words and term.

IV. PROPOSED SYSTEM SUMMARY

In this proposed approach, we introduce an algorithm for the automatic extraction (or validation) of policy networks using information collected from the web. Specifically, 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. The proposed algorithm is 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. A policy network can be considered as a type of social graph in which the nodes represent the actors involved in a given policy field, while their relations are represented by the edges. In social networks, nodes usually correspond to persons and edges represent relations among them built on a ground of mutual understanding, which can take several forms, such as friendship and coauthor ship. In policy networks, actors can be organizations or even groups or unions of variable size and degree of formal organization.

The relations among such actors usually signify the development of partnerships rather than a lax social relation. Furthermore, relations in policy networks depend on external factors, such as economic policies and funding at the local, national or supra-national level. Relations between policy actors (much like economic actors) can also be antagonistic rather than cooperative, or follow a more complex pattern of both cooperation and competition (sometimes referred to by economists as competition). Often policy networks are studied at their infancy when the links between actors are just emerging and might not be directly observable through common action or direct communication. All these subtleties imply that established features and algorithms for social network analysis might not be directly applicable to policy network extraction.

V. LITERATURE SURVEY

In the year of 2012, the authors " H. Gao, J. Hu, C. Wilson, Z. Li, Y. Chen" proposed a paper titled "Detecting And Characterizing Social Spam Campaigns", in that they described such as: Online social networks (OSNs) are popular collaboration and communication tools for millions of users and their friends. Unfortunately, in the wrong hands, they are also effective tools for executing spam campaigns and spreading malware. Intuitively, a user is more

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

likely to respond to a message from a Facebook friend than from a stranger, thus making social spam a more effective distribution mechanism than traditional email. In fact, existing evidence shows malicious entities are already attempting to compromise OSN account credentials to support these "high-return" spam campaigns. In this paper, we present an initial study to quantify and characterize spam campaigns launched using accounts on online social networks. We study a large anonymized dataset of asynchronous "wall" messages between Facebook users. We analyze all wall messages received by roughly 3.5 million Facebook users (more than 187 million messages in all), and use a set of automated techniques to detect and characterize coordinated spam campaigns. Our system detected roughly 200,000 malicious wall posts with embed- ded URLs, originating from more than 57,000 user accounts. We find that more than 70% of all malicious wall posts advertise phishing sites. We also study the characteristics of malicious accounts, and see that more than 97% are compromised accounts, rather than "fake" accounts created solely for the purpose of spamming. Finally, we observe that, when adjusted to the local time of the sender, spamming dominates actual wall post activity in the early morning hours, when normal users are asleep.

In the year of 2013, the authors "H. Gao, Y. Chen, K. Lee, D. Palsetia" proposed a paper titled "Towards Online Spam Filtering In Social Networks", in that they described such as: Social network has become a very popular way for internet users to communicate and interact online. Users spend plenty of time on famous social networks (e.g., Facebook, Twitter, Sina Weibo, etc.), reading news, discussing events and posting messages. Unfortunately, this popularity also attracts a significant amount of spammers who continuously expose malicious behavior (e.g., post messages containing commercial URLs, following a larger amount of users, etc.), leading to great misunderstanding and inconvenience on users′ social activities. In this paper, a supervised machine learning based solution is proposed for an effective spammer detection. The main procedure of the work is: first, collect a dataset from Sina Weibo including 30,116 users and more than 16 million messages. Then, construct a labeled dataset of users and manually classify users into spammers and non-spammers. Afterwards, extract a set of feature from message content and users′ social behavior, and apply into SVM (Support Vector Machines) based spammer detection algorithm. The experiment shows that the proposed solution is capable to provide excellent performance with true positive rate of spammers and non-spammers reaching 99.1% and 99.9% respectively.

In the year of 2014, the authors "P. Chia, Y. Yamamoto, And N. Asokan" proposed a paper titled "This App Safe? A Large Scale Study On Application Permissions And Risk Signals", in that they described such as: Third-party applications (apps) drive the attractiveness of web and mobile application platforms. Many of these platforms adopt a decentralized control strategy, relying on explicit user consent for granting permissions that the apps request. Users have to rely primarily on community ratings as the signals to identify the potentially harmful and inappropriate apps even though community ratings typically reflect opinions about perceived functionality or performance rather than about risks. With the arrival of HTML5 web apps, such user-consent permission systems will become more widespread. We study the effectiveness of user-consent permission systems through a large scale data collection of Facebook apps, Chrome extensions and Android apps. Our analysis confirms that the current forms of community ratings used in app markets today are not reliable indicators of privacy risks of an app. We find some evidence indicating attempts to mislead or entice users into granting permissions: free applications and applications with mature content request more permissions than is typical; 'look-alike' applications which have names similar to popular applications also request more permissions than is typical. We also find that across all three platforms popular applications request more permissions than average.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

VI. EXPERIMENTAL RESULTS

The following figure shows the Home Page of the proposed system.

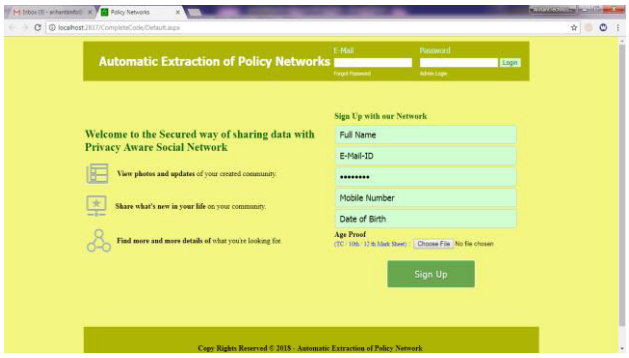

Fig.1 Home Page

The following figure illustrates the Administrator Login Page of the proposed system.

Fig.2 Administrator Authentication

The following figure illustrates the view of registered users.

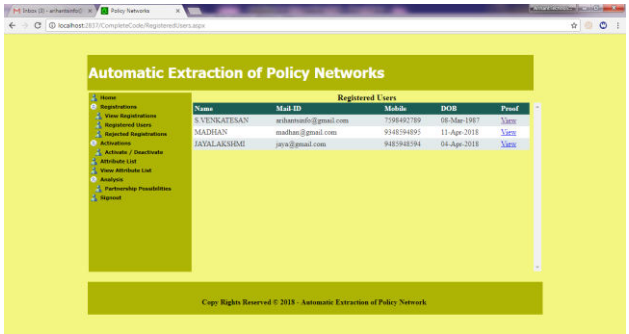

Fig.3 Registered Users

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

The following figure illustrates the partnership possibilities of the proposed system.

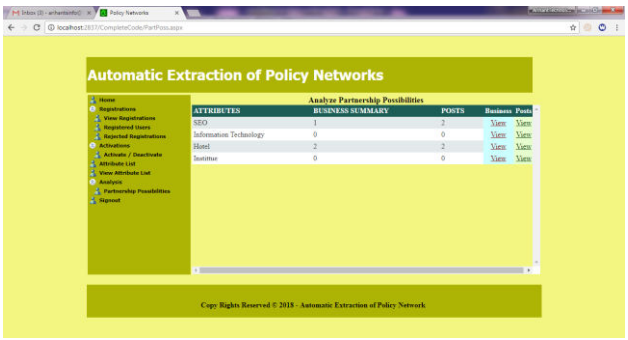

Fig.4 Partnership Possibilities

VII. CONCLUSION

The proposed system is higher efficiency than the existing system. System is identifying automatically the actors participating in policy networks and their lexicalizations. And also in the proposed system, the semantic ranking function is introduced to estimate the semantic similarity higher. It is filtering web data based on relevance and type of source for instance nonprofit, government, corporate, blogs and news. Then apply the machine learning algorithms such as SVM to select the most informative metric for the additional features. Then apply the machine learning algorithms such as SVM to select the most informative metric for the additional features. In the proposed system, we apply aggregate index search and twofold search approaches for computing nearest neighbor spatial and social proximities in social networks. Also we enhance the graph search distance for finding shortest path distance in the given social network. Thus the proposed system methods are useful to retrieve the more semantic social information. The experimental result proves that the proposed system has highest performance rather than the existing system.

REFERENCES

[1] Tanja A.Borzel, "An exploration of the concept and its usefulness in studying European governance," European Integration online papers, vol. 1, pp. 016.
[2] R.A.W. Rhodes,"Policy network analysis", John Donne [1611], 'The First Anniversary. An Anatomy of the World', 1985 edition, 335 line 213.
[3] Theodosis Moschoulos,Elias Iosif, "Towards automatic extractiojn of policy networks usinf web links and documents," in IEEE, vol. 25, No. 25, 2013, pp. 2403-2417.
[4] Kyriakos Mouratidis, Jing Li, Yu Tang and Nikos Mamoulis, "Joint search by social and spatial proximity," in IEEE transactions on knowledge and data engineering, Vol 27, No. 3 March 2015.
[5] Burt L.Monroe, "introduction to the special issue: The statistical analysis of political text,"in Advance access publication, February 4, 2009.
[6] L.Zhu, Computational Political Science Literature Survey, http://www.personal.psu.edu/luz113, 2013.
[7] Michael laver and Kenneth benoit, "Extracting policy positions from political texts using words as data,"American political science review, vol. 97, No. 2,May 2003
[8] Jonathan B. Slapin and Sven-oliver Proksch,"A scling model for estimating time-series party positions from texts" Trinity college, Dublin and University of california, Los Angeles.
[9] Nick Craswell, David Hawking and Stephen Robertson " Effective site finding using link anchor information" .
[10] Yutaka Mastsuo, Junichiro Mori and Masahiro Hamasaki " POLYPHONET: An advanced social network extraction system from the web" in National institude of Advanced industrial science and technology, United Staes, University of Tokyo, Japan.
[11] D. Bollegala, Y. Matsuo, and M. Ishizuka, "Measuring Semantic Similarity between Words Using Web Search Engines," Proc. 16th Int'l World Wide Web Conf., pp. 757-766, 2007.
[12] R. Cilibrasi and P. Vitanyi, "The Google Similarity Distance," IEEE Trans. Knowledge and Data Eng., vol. 19, no. 3, pp. 370-383, Mar. 2007.
[13] P. Vitanyi, "Universal Similarity," Proc. Information Theory Workshop Coding and Complexity, pp. 238-243, 2005.
[14] Lourdes Araujo and Juan Martinez- Romo, " Web Spam Detection: Mew Classification Features Based on Qualified Link Analysis and Language Model" in IEEE Trans. On Information Forensics and Security, Vol. 5, No. 3 September 2010.
[15] George Tsatsaronis and Iraklis Varlamis, " Semantic Rank: Ranking Keywords and Sentences using Semantic Graphs, in Norwegian University of Science and technology
[16] Priyanka N Guttedar and Pushpalata S, " Scene Text Recognition in Mobile Application using K- Mean Clustering and Supporrt Vector Machine", In IJARCET, Vol. 4 5 May 2015.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 4, April 2018

[17] L. Adamic, O. Buyukkokten, E. Adar, "Social network caught in the Web",First Monday 8 (6).
[18] J. Golbeck, J. Hendler, Accuracy of metrics for inferring trust and reputationin semantic Web-based social networks, in: Proceedings of the EKAW 2004, 2004.
[19] P. Mika, Flink, Semantic Web technology for the extraction and analysis of social networks, J. Web Semantics 3 (2) (2005) 211–223.
[20] M. Harada, S. Sato, K. Kazama, Finding authoritative people from theWeb,in: Proceedings of the Joint Conference on Digital Libraries (JCDL2004),2004.
[21] P. Turney, Mining theWeb for synonyms: F PMI-IR versusLSAon TOEFL,pp. 491–502, in: Proceedings of the ECML-2001, 2001.
[22] J. Goecks, E.D. Mynatt, Leveraging social networks for information sharing, pp. 328–331, in: Proceedings of the ACM CSCW, 2004.