Government Agency Network Discovery via Hyperlinks and Datalinks

Li Ding[1], Ryan Whalen[2], Deborah L. McGuinness[1], Noshir Contractor[2]
[1] Tetherless World Constellation, Rensselaer Polytechnic Institute, 110 8th St. Troy, NY 12180, USA
[2] SONIC Lab, Northwestern University, Evanston, IL 60208, USA
dingl@cs.rpi.edu, ryan@ryanwhalen.com, dlm@cs.rpi.edu, nosh@northwestern.edu

ABSTRACT
In order to improve government transparency and promote interagency coordination, interagency networks have been investigated by social science researchers using data collected using traditional manual mechanisms such as interview and survey. In this paper, we 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. The outcome of this study includes two independent interagency networks connecting US Federal government agencies, unveiling several interesting interagency connections which have never been seen before.

Keywords
Data.gov, Linked Data, interagency network, open government

1. INTRODUCTION
USA.gov provides a comprehensive list of US federal government agencies and a hierarchical organization chart (see Figure 1) illustrating the administrative relations among the executive, legislative and judicial branches. While this provides a lot of useful information for citizens to explore individual government agencies, there exists little, if any, information about how these federal agencies are inter-connected and how they cooperate.

Apart from their hierarchical administrative organization, US federal government agencies maintain dynamic public web presences, allowing further investigations into the structural patterns among agencies based on the underlying World Wide Web (WWW) graph where agencies are connected hyperlink paths as well as the Linked Open Data (LOD) graph where agencies are connected via various semantic relations. An important feature of such connections is that they can be automatically collected from the existing Web infrastructure, complementing the traditional manual data collection process such as interviews and surveys. In this paper we leverage the Web as the primary source to produce two contributions:

- This work shows that interagency networks can be extracted from Web data automatically using different methods, including hyperlink analysis and linked data analysis.
- The resulting interagency networks reveal novel insights about the structuring of US federal government agencies. The results are useful to support government transparency as well as facilitate interagency cooperation.

WebSci '11, June 14-17, 2011, Koblenz, Germany.

2. RELATED WORK
Network analysis has been widely used in social science to discern structure (Borgatti et al. 2009, Monge & Contractor, 2003). Within government domain, these include *inter-organizational networks* and *inter-personal networks*.

Oliver (1990) analyzed six types of inter-organizational networks (via e.g. financial relations and political cooperation) and discussed the determinants that lead to the formation of these networks. Krauss et al. (2004) explored state tobacco control inter-agency networks from US states including New York, Michigan, Indiana, Wyoming and Washington. Recently, White (2008) constructed an inter-agency network based on organizations' partnership relations to facilitate network learning and growth. This line of work provides interesting results of inter-agency network analysis, but most results are limited to small scales because the data are collected using traditional manual data collection methods such as questionnaires, surveys, interviews and workshops.

Similarly, interpersonal networks have been constructed and analyzed, e.g. municipal managers' "co-membership of regional associations and professional associations" (LeRoux et al. 2010), and congressmen's "co-sponsorship of congressional legislation" (Cho and Fowler 2010). While the latter study starts leveraging an existing dataset, i.e. co-sponsorship network dataset generated by Fowler (2006), the former uses data collected from interviews.

To the best of our knowledge, few, if any, studies have used the Web as a knowledge source to discern inter-organizational networks in the government domain specifically, the US federal government agencies.

3. INTERAGENCY NETWORKS
Within the traditional organizational depiction of the US Federal government, all department level agencies are treated as equals and only administrative relationships are described in the hierarchy (see Figure 1). This traditional structural representation omits lateral inter-agency relationships. The web presence of these agencies is indeed interconnected at multiple levels, via hyperlinks and linked data. In what follows, we show two use-cases on how Web data can be used to extract the latent semantic relationships among these agencies.

3.1 Hyperlink-based Agency Network

The .gov domain is home to the US federal government's web presence: US federal agencies typically maintain unique domain names such as usda.gov or epa.gov. These government agency websites are interconnected via one-hop (or multi-hop) directed hyperlink paths. Analyzing the hyperlink connections between these government websites allows us to reveal relationships between government agencies that are not otherwise evident.

Utilizing Lexiurl (http://lexiurl.wlv.ac.uk/) - software designed to access *Yahoo!* web graph information - this study started at USA.gov, the hub listing all federal agencies' home websites, and determined the linking relationships between federal websites, building a network of domain names (each of which maps to a unique federal government agency). More details about the process can be found in (Whalen 2011).

Figure 2 shows an example of agency networks connecting department-level agencies, demonstrating that some agencies (i.e. Treasury, Commerce) are more peripheral to the agency network while others (i.e. Labor, USDA) are more centrally situated. It should be noted that the nodes in this network are government agencies and the links indicate that the web site of one agency has a link that points to the web site of another government agency.

3.2 Datalink-based Agency Network

Agencies can also be connected by shared interests. For instance, the common phrases found in the description of two agencies can be a good indicator of common interests of the two agencies. As a result of campaigns for "putting government data online" (Berners-Lee 2009), a number of open government datasets owned by different agencies have now been published on the Web. Therefore we can now connect government agencies by the commonality between their owned datasets. Here, mutual datalinks (e.g. sharing common interests and running joint programs) between agencies can be discovered by analyzing linked data that connects datasets to individual agencies. Note that datalinks differs from hyperlinks as it could carry different semantics and is derived from correlation analysis on linked data, i.e. it does not depends on hyperlinks' web addressing semantics.

Our experiment reuses the results of our resent Linking Open Government Data (LOGD) project (Ding et al. 2010), where we convert, enhance and mash-up datasets in Data.gov, the official web portal for US open government data (OGD), into linked open government data. In particular, we used Data.gov's dataset catalog to extract interagency datalinks, i.e. agencies are connected if their published relevant/similar datasets on Data.gov.

As shown in Figure 3, Department of Energy connects with a couple of other peer agencies, and the hierarchical relation between Department of the Interior and EPA is also reflected in datalinks. By analyzing the topics of datasets, we can further explain how the agencies are connected. A deeper analysis on the related datasets shows that DOE connect to National Science Foundation via DOE's science acceleration widget, to Department of Labor via coal production, and to NASA via solar energy.

4. CONCLUSIONS

Running a novel multi-modal network creation and analysis on information collected from the Web, this study provides insight into the structure of the federal government unavailable in traditional organizational analyses. This work complements the traditionally costly and limited manual data collection methods such as interview and survey. The agency networks discovered by this study also (i) help citizens understand how different agencies collaboratively approach their e-government missions on the Web, and (ii) unveil the seemingly "opaque" organizational chart with factual and meaningful indicators from the Web. Future work will apply these findings to supplement subsequent organizational studies on e-government and to assist in real-world governmental interagency network planning and growth.

Acknowledgement: this work is partially supported by the following funding sources: Army Research Laboratory under Cooperative Agreement Number W911NF-09-2-0053, National Science Foundation Grants: CNS-1010904, OCI-0904356, IIS-0838564, DARPA's Transformational Convergence Technology Office, and gifts from Lockheed Martin, Fujitsu Labs America, LGS, and Microsoft.

5. REFERENCES

[1] Borgatti, S.P., Mehra, A., Brass, D. J. and Labianca, G. (2009), Network Analysis in the Social Sciences. *Science*, February 2009, 323 (5916), 892-895

[2] Monge, P. R. and Contractor, N. (2003), *Theories of Communication Networks*. Oxford University Press

[3] Christine Oliver (1990), Determinants of Interorganizational relationships: Integration and Future Directions. *Academy of Management Review*, April 1990, 15(2), 241-265

[4] Krauss, M., Mueller, N. and Luke, D. (2004), Interorganizational relationships within state tobacco control networks: a social network analysis. *Preventing chronic disease*, October 2004, 1(4), A08.1

[5] White, L. (2008), Connecting organizations: Developing the idea of network learning in inter-organizational settings, *Systems Research and Behavioral Science*, November 2008, 25(6), 701–716

[6] Cho, W. K. T. and Fowler, J. H. (2010), Legislative Success in a Small World: Social Network Analysis and the Dynamics of Congressional Legislation. *Journal of Politics*, January 2010, 72(1), 124–135

[7] LeRoux, K., Brandenburger, P. W. and Pandey, S. K. (2010), Interlocal Service Cooperation in U.S. Cities: A Social Network Explanation. *Public Administration Review*, March/April 2010, 70(2), 268–278

[8] Fowler, J. H. (2006), Connecting the Congress: A Study of Cosponsorship Networks. *Political Analysis*, Fall 2006, 14(4): 456-487

[9] Whalen, R. (2011) The Structure of Federal eGovernment: Using Hyperlinks to Analyze The .Gov Domain. Proceedings of Political Networks Conference, Ann Arbor, MI, 2011

[10] Berners-Lee, T. (2009). Putting government data online. http://www.w3.org/DesignIssues/GovData.html

[11] Ding, L., Difranzo, D., Graves, A.. Michaelis, J., Li, X., McGuinness, D. L., Hendler, J. (2010), TWC data-gov corpus: Incrementally generating linked government data from data.gov, *Proceedings of the 19th International World Wide Web confere*nce *(WWW2010) (developer track)*, 2010.

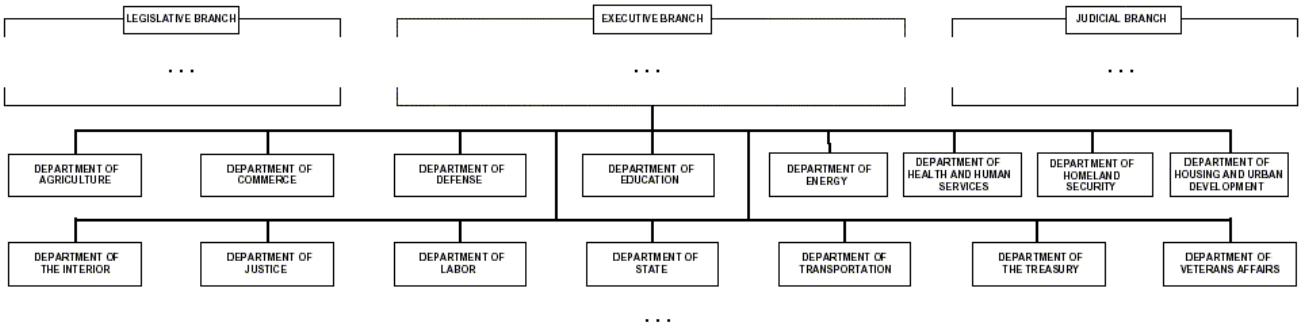

Figure 1. The hierarchical organization of Federal Government agencies (source: USA.gov)

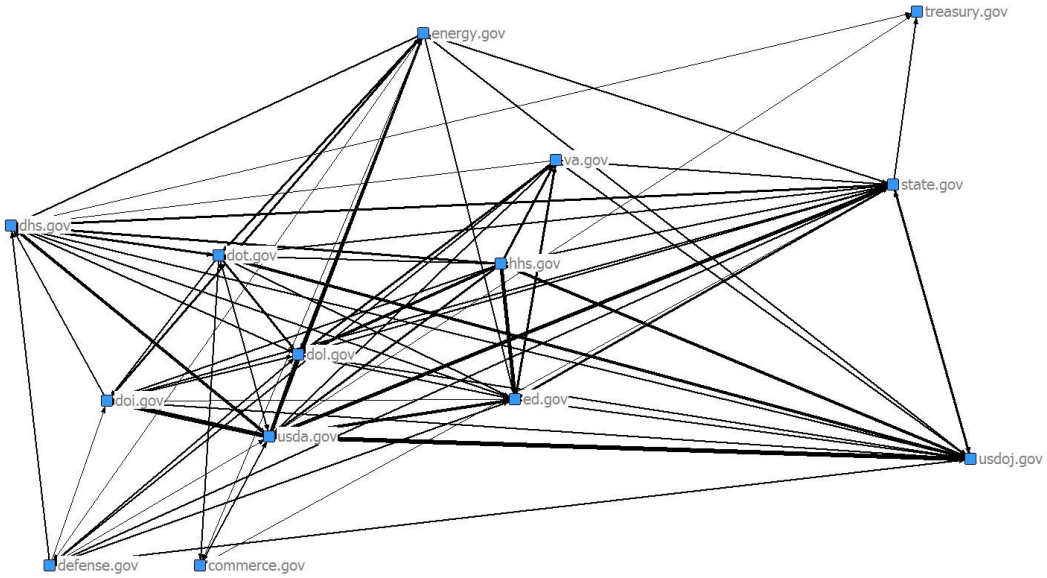

Figure 2. A fragment of the agency network derived from hyperlinks.

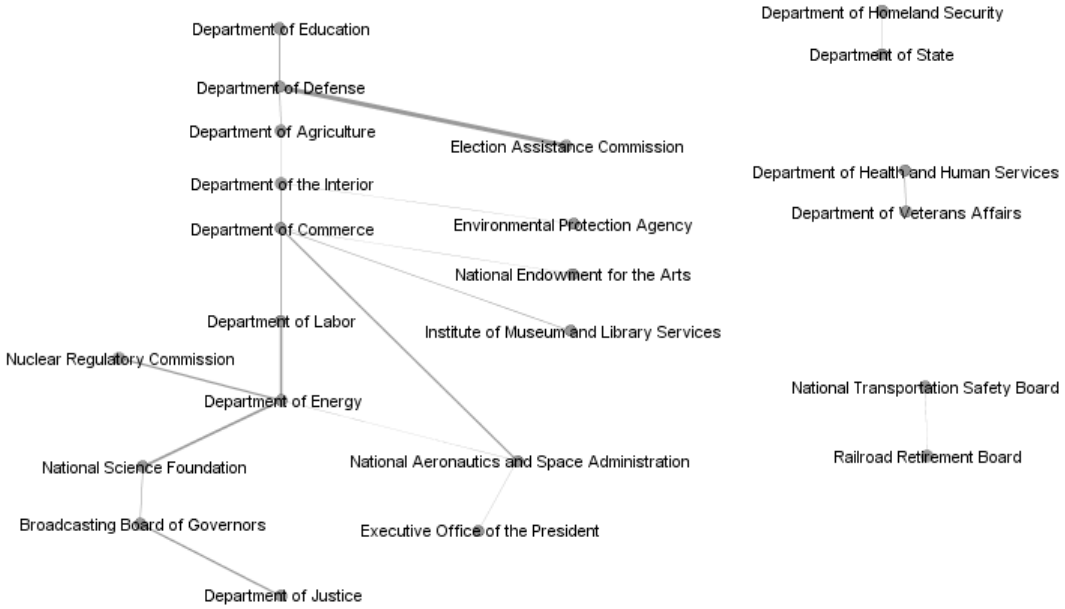

Figure 3. A fragment of agency network derived from datalinks