

From Media Reporting to International Relations: A Case Study of Asia-Pacific Economic Cooperation (APEC)

Chun-Hua Tsai
School of Information Sciences
University of Pittsburgh
Pittsburgh, PA 15260, USA
cht77@pitt.edu

Yu-Ru Lin
School of Information Sciences
University of Pittsburgh
Pittsburgh, PA 15260, USA
yurulin@pitt.edu

ABSTRACT

Network analysis has brought new perspectives in studying emergent structure in international relations. Prior work on networks of international relations has mostly relied on data derived from formal alliances and trade flows, which can hardly capture the rapidly evolving international relations due to globalization and recent advances in information technology. In this work, we propose a novel research design that aims at capturing “real-time” international relations through news reporting. We collect worldwide news on a daily basis, and characterize the relationship between any two countries through analyzing the similarity of their news content. Our empirical results based on news about the APEC (Asia-Pacific Economic Cooperation) CEO Summit in 2013 revealed interesting and meaningful international relations among member countries.

Categories and Subject Descriptors

H.2.8 [Database Management]: Database applications—*Data mining*; H.4 [Information Systems Applications]: Miscellaneous

General Terms

Measurement; Experimentation

Keywords

online news media, international relations, network analysis

1. INTRODUCTION

The rapid development in technologies, such as transportation and telecommunication infrastructure, has expedited the process of globalization at almost every corner in the world. This globalization process has distinguished speed and volume of integration of political, economic and cultural activities across country borders, through the exchanges of views, ideas, labors, products and other aspects

of culture, which permeates everyone’s life. How can we capture such ever-changing cross-country relations in the real-time age?

Most of the prior work on networks of international relations has relied on data derived from formal alliances and trade flows [3]. These studies have provided important insights to the understanding of historic and contemporary international relations covering political and economic aspects. Nevertheless, the data employed in these studies are insufficient to capture the up-to-date, rapidly evolving international relations because of two reasons: the data are collected with considerable time lag, and the datasets often do not cover the diverse and fluid views or ideas possessed by different cultures.

In this work-in-progress paper, we present a novel research design that aims at providing a research infrastructure to capture “real-time” international relations through news reporting. Our design is inspired by the observation that news from different countries reflects different cognitive representations of the world [6]. We develop a system to collect the news articles from Google News website on a daily basis. We then apply text analysis, including matrix factorization based document clustering technique, on the news corpus. Based on the topical similarity between countries’ reporting news, we construct time-varying networks among countries that capture the cross-country relations through their common or different topical interests attended in the daily news.

2. APPROACH

We propose a system to collect the news from Google News website into our database. Google News is an automatic-aggregated news website that provides news from thousands news outlets in 70 regional and countries with 20 languages [1]. This is one of the sources that allow for collecting daily news with worldwide coverage. In our system, we implement a web spider to monitor the Google News by different countries, at an hourly basis. This system collects and parses texts from HTML source pages after downloading the news pages. In order to analyze these the global news articles, we employ the Google Translate function to convert non-English texts into English texts [5]. We then remove the stop words and perform stemming on the translated English texts. We then create term-document matrix for further analyses.

We apply Non-negative Matrix Factorization (NMF) on the term-document matrix to discover countries’ similar topical interests in news reporting as “themes” [2, 4]. NMF is a matrix factorization technique that is used to transform

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

WebSci’14, June 23–26, 2014, Bloomington, IN, USA.
2014 ACM 978-1-4503-2622-3/14/06.

the term-document matrix into word-themes and theme-documents matrix. The non-negative feature provides the better interpretability from the text [4]. Based on the theme-document matrix obtained from NMF and the documents' source countries, we create two types of networks. (1) *Country-to-theme network* is a bipartite network where connecting two types of nodes, country and theme nodes, and edge weights are computed by aggregating all documents from the same countries. (2) *Country-to-country network* is a unipartite network where nodes are countries, and the weight of an edge between two countries is given by the cosine similarity of the two countries' aggregated theme distributions.

3. CASE STUDY

We use APEC CEO Summit 2013 as our case study. This meeting was held in Bali, Indonesia from 05 Oct, 2013 to 07 Oct, 2013. There were 21 member countries participated in this meeting. Four countries' news (Indonesia, Thailand, Brunei Darussalam and Papua New Guinea) are missing by Google News. Hence, we only collected 17 member countries' news articles in our dataset. The news volume peaked at Oct. 07. In our analysis, we focus on news that contains the keyword "APEC" during the two weeks centered on Oct. 07 (from Oct. 1 to 15, 2013). Figure 1(a)(b) show the news frequency of APEC.

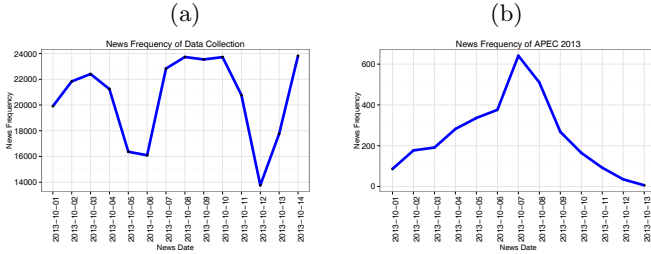


Figure 1: (a) News Frequency of Data Collection. (b) News Frequency of APEC 2013.

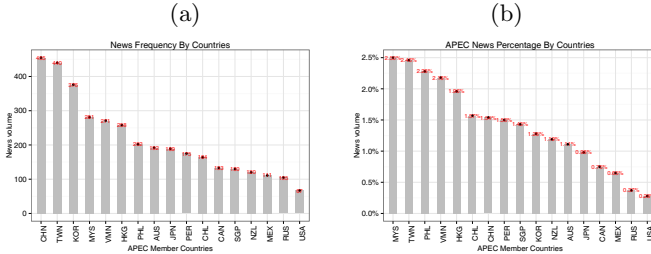


Figure 2: (a) News frequency by countries: China and US had the highest and lowest APEC news volumes, respectively. US has relative low volume of news articles due to the absence of US president.

4. RESULTS

We tested the number of themes from 2 to 10 in NMF. Six themes are selected based on manual inspection as they covers the major events well. The six themes are: C1) Trade and economic growth issues; C2) China president Xi Jin Ping; C3) the cross-strait relation between Taiwan and China; C4) United Government shutdown and President Obama absence; C5) Association of Southeast Asian

Nations and South Sea dispute; C6) The conflict between Hong Kong and Philippine.

Figure 3(a) and (b) present the networks of country-to-theme and country-to-country, respectively. In this international event, most countries mentioned about trade and growth issues. China's news reporting covered a diverse set of issues, while other countries had relatively limited interests. Taiwan is the only country focusing on the cross-strait relation. US had relatively weak connection with other countries due to lower media attention about this meeting, which was probably due to Obama's absence.

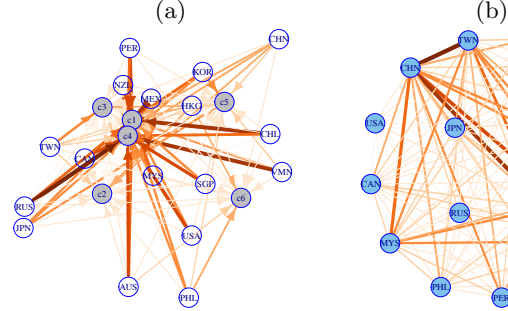


Figure 3: (a) Country-to-theme network: Theme C1 and C4 are dominant themes; (b) Country-to-country network: China, Japan and Korea have strong similarity in news reporting.

5. CONCLUSION

We have experimented our framework on news articles about the APEC CEO Summit held Indonesia 2013. Our empirical results suggested that the content similarity between the participating countries' new articles revealed the structure captured by international trade flows, but also capture emergent relationships and roles in these countries. Our initial finding also has important implications for analyzing online media corpus for understanding underlying social and political changes on a global scale.

6. REFERENCES

- [1] A. S. Das, M. Datar, A. Garg, and S. Rajaram. Google news personalization: scalable online collaborative filtering. In *Proceedings of the 16th international conference on World Wide Web*, pages 271–280. ACM.
- [2] R. Gaujoux and C. Seoighe. A flexible r package for nonnegative matrix factorization. 11(1):367.
- [3] E. M. Hafner-Burton, M. Kahler, and A. H. Montgomery. Network analysis for international relations. 63(3):559–592.
- [4] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. 401(6755):788–791.
- [5] J. Savoy and L. Dolamic. How effective is google's translation service in search? 52(10):139–143.
- [6] E. Segev and T. Hills. When news and memory come apart: A cross-national comparison of countries mentions. 76(1):67–85.