Chapter 9
Exploring "User," "Video," and (Pseudo) Multi–Mode Networks on YouTube with NodeXL

Shalin Hai-Jew
Kansas State University, USA

ABSTRACT

Network Overview, Discovery and Exploration for Excel (NodeXL Basic) enables the extraction of "user" (entity), "video" (content), and pseudo multi-modal networks from YouTube. This open-source add-on to NodeXL captures a wide range of data, enables data processing for analysis, and then visualization in a variety of graphs (based on different layout algorithms). This chapter summarizes some of the "ask-able" questions using this approach. Various types of data extractions are shared to give a sense of the breadth of some approaches, including the following: (1) entities, (2) in-world phenomena, (3) imaginary phenomena, (4) themes, (5) reputations by name, (6) genres, (7) language-specific phenomena, and (8) location-specific phenomena.

INTRODUCTION

The work of this chapter overlapped with the ten-year anniversary of the start of YouTube. In the lore, the first YouTube video ever posted was titled "Me at the zoo" (by YouTube co-founder Jawed Karim); in the decade since, the historic video "in praise of elephants" as racked up almost 20 million views (Newcomb, Apr. 23, 2015). As a social media platform, YouTube is a formidable and dominant player, with billions of video views daily and presences across the globe. YouTube is "the world's second-largest search engine next to Google" (Rogers & Krishnan, 2014, p. 84). The functioning of the site depends in part on the collection of various text-based metadata about the videos, and the collection of user- and video- relationship data.

This metadata and trace data may be extracted and visualized through NodeXL (Network Overview, Discovery and Exploration for Excel), a free and open-source add-on to Excel. The structuring of user networks, video networks, and two-mode / multi-mode networks, enables insights about the users of the

social media platform, the video contents, and other information that would otherwise be latent or hidden. Various types of data extractions (based on topics that were contemporaneous at the time of the data extraction) are shared here to give a sense of the breadth of some approaches. This chapter will address some of the types of questions that may be asked using this research tool and some of the embedded methodologies. Real-world examples using empirical data were used and include some of the following types of seeding contents to extract the data and networks from YouTube, including the following:

1. Entities,
2. In-world phenomena,
3. Imaginary phenomena,
4. Themes,
5. Reputations by name,
6. Genres,
7. Language-specific phenomena, and
8. Location-specific phenomena.

This list is a partial one only and is meant to be used to potentially inspire reader ideas for research.

This chapter includes a selective summary of some of the academic research about YouTube, examples of "video," "user," and two-mode networks extracted from YouTube, some of the "askable" questions using this software add-on, and an overview of the two YouTube-based data extraction and modeling features of the NodeXL tool.

One Caveat about the Data Visualizations

The data visualizations here include the use of some colors that may be less readable when processed in black-and-white and in paper format. The electronic versions of this chapter will likely read much more clearly than the print version. The electronic version would also enable zooming in for close-in analysis.

A SELECTIVE REVIEW OF THE LITERATURE

Founded in February 2005 and purchased by Google in October 2006, YouTube was designed initially as a platform for people to broadcast themselves. Since then, it has become the premiere video-sharing site in the world, with localization in 75 countries and availability in 61 languages. Some 300 hours of video are uploaded to YouTube every minute (Newcomb, Apr. 23, 2015). This video-sharing platform itself is a sophisticated one. It has an automated and built-in system to detect contravening of copyright or intellectual property rights—in text, audio, visual, video, and multimedia forms. [Its Content ID program enables the protection of copyright for text, audio, and video—and the company has paid out $1 billion since 2007 to those who monetized their copyright claims ("Statistics," 2015)]. Since 2009, it has integrated a built-in machine-based speech-recognition tool to enable audio-to-text annotation (in ten different languages), which works in a complementary way to human corrected transcription for timed text captioning. YouTube uses owner-uploaded video transcripts to generate additional semi-supervised training data and deep neural networks acoustic models with large state inventories, according to Google researchers (Liao, McDermott, & Senior, 2013, p. 368).

Related Content

Emotion-Drive Interpretable Fake News Detection
Xiaoyi Ge, Mingshu Zhang, Xu An Wang, Jia Liuand Bin Wei (2022). *International Journal of Data Warehousing and Mining (pp. 1-17).*
www.irma-international.org/article/emotion-drive-interpretable-fake-news-detection/314585

Image Mining: Detecting Deforestation Patterns Through Satellites
Marcelino Pereira dos Santos Silva, Gilberto Câmaraand Maria Isabel Sobral Escada (2009). *Data Mining Applications for Empowering Knowledge Societies (pp. 55-75).*
www.irma-international.org/chapter/image-mining-detecting-deforestation-patterns/7546

Multidimensional Design Methods for Data Warehousing
Oscar Romeroand Alberto Abelló (2011). *Integrations of Data Warehousing, Data Mining and Database Technologies: Innovative Approaches (pp. 78-105).*
www.irma-international.org/chapter/multidimensional-design-methods-data-warehousing/53073

Coupling Materialized View Selection to Multi Query Optimization: Hyper Graph Approach
Ahcene Boukorca, Ladjel Bellatreche, Sid-Ahmed Benali Senouciand Zoé Faget (2015). *International Journal of Data Warehousing and Mining (pp. 62-84).*
www.irma-international.org/article/coupling-materialized-view-selection-to-multi-query-optimization/125651

On Nonredundant Cost-Constrained Team Formation
Yu Zhou, Jianbin Huang, Heli Sunand Xiaolin Jia (2017). *International Journal of Data Warehousing and Mining (pp. 25-46).*
www.irma-international.org/article/on-nonredundant-cost-constrained-team-formation/185657