Chapter 15
Spatiotemporal Network Analysis and Visualization

Judith Gelernter
Carnegie Mellon University, USA

Kathleen M. Carley
Carnegie Mellon University, USA

ABSTRACT

Spatiotemporal social network analysis shows relationships among people at a particular time and location. This paper presents an algorithm that mines text for person and location words and creates connections among words. It shows how this algorithm output, when chunked by time intervals, may be visualized by third-party social network analysis software in the form of standard network pin diagrams or geographic maps. Its data sample comes from newspaper articles concerning the 2006 Darfur crisis in Sudan. Given an immense data sample, it would be possible to use the algorithm to detect trends that would predict the next geographic center(s) of influence and types of actors (foreign dignitaries or domestic leaders, for example). This algorithm should be widely generalizable to many text domains as long as the external resources are modified accordingly.

1. INTRODUCTION

1.1. Social Networks and Assumptions about Place and Time

A social network describes a group of people who share some sort of social connections, whether through work, or friendship, or otherwise. The social network concept stems from mid 20th century sociology. Alignment of social network studies with computer science in about the 1970s allowed the connections among individuals to be weighted and computed mathematically on a large scale, with weights indicating, for example, strength of the relationship.

An analysis of a social network generally focuses on the groupings of people. The people might be employees of an organization, for example, or colleagues in a discipline, sportsmen on teams, or characters in a novel. Questions that could be answered by the analysis include: Which individuals are

DOI: 10.4018/978-1-4666-9845-1.ch015

in what group? Who leads each group? Who is second to the leader? Who is between two groups? Are group relations friendly or antagonistic? At the foundation of the social network literature are Scott (1991 [2004]), and Wasserman and Faust (2004).

Social networks have been diagrammed with people as points and their social relations as lines. The points are called nodes; the relations are called ties or edges. The groups are cliques. The aggregate of cliques form a network at some time. Most social network studies and the standard social network diagrams account for neither space nor time. This may be because it is assumed that the network is spatially and temporally persistent and so does not change, or because a network snapshot is good enough. Whichever assumption is made, often space and time are treated as irrelevant factors.

Our social networks in our research are formed from connections among people mentioned in news articles. The activities of people and their co-relationships in space and time come from the context of the news story. When we extract names of peoples from these texts and build links among them using the proximity of their names in the article, we are in essence attributing a relationship among the people. This context can be characterized by the spatio-temporal setting of the news article.

Social network analysis of vast amounts of text through data mining, as described in this research, affords an overview of events. No reading of the text is necessary. This does not simply save an analyst much time and effort; it allows assimilation of text on a scale that would otherwise require many people to analyze. Even though errors occur because network nodes are only *inferred,* automated node extraction saves times and offers insight that is valuable. We suggest that the accuracy of the extracted network can be enhanced through improved extraction of spatio-temporal information that describes the network membership and relations among members.

1.2. Data Mining For Social Networks

Data mining techniques are used to extract social network data from text. The way it works is that the text is submitted to a series of filters until the sought-after information remains. In the early phases of processing, grammatical articles (a, an, the) are filtered as noise. Sometimes numerals and symbols in the text are removed, and the text is normalized to lower case. Remaining words may be reduced to their stems so that noun plurals, verb past tenses, gerund endings and so forth are removed. Then external lists will be more effective in finding relevant network words in the text.

How do we mine for people and location nodes? To identify person and place, external sources as well as language processing methods play a role. In Named Entity Recognition, an entity is a proper name, an organization, an event, or a location (Giuliano, Lavelli, & Romano, 2007). To simplify the data mining required in our case, we restricted the entities to names on a match list, and to places in a world gazetteer. We mine date from the header information that accompanies each news article. A bibliography for the mining of network data for spatio-temporal information is found in Roddick and Spiliopoulou (1999).

How do we determine edges that are between nodes? Edges are created according to node word proximity as dictated by grammar and syntax, and within a word-window size set by the text-processing software user. As an example, person node and location node found anywhere in the same sentence ordinarily will receive an edge. If that same person node ends one sentence and the location begins the following sentence, they might not receive an edge if the proximity window size set by the user had been small, although they would receive an edge if the proximity window set had been large. Thus, the process of setting node edges differs widely. Edges are determined irrespective of the meaning of the sentence from which they are extracted. Extracting network edges that are meaningful is substantially

Related Content

Fire Recurrence and the Dynamics of the Enhanced Vegetation Index in a Mediterranean Ecosystem

Dania Abdul Malak, Juli G. Pausas, Josep E. Pardo-Pascualand Luis A. Ruiz (2019). *Geospatial Intelligence: Concepts, Methodologies, Tools, and Applications (pp. 1690-1708).*

www.irma-international.org/chapter/fire-recurrence-and-the-dynamics-of-the-enhanced-vegetation-index-in-a-mediterranean-ecosystem/222968

Forest Cover Change in the Northeastern U.S.: A Spatial Assessment in the Context of an Environmental Kuznets Curve

George C. Bentley, Robert G. Cromley, Dean M. Haninkand C. Patrick Heidkamp (2013). *International Journal of Applied Geospatial Research (pp. 1-18).*

www.irma-international.org/article/forest-cover-change-northeastern/77922

Building Information Modelling (BIM) for Facilities Management (FM): The Mediacity Case Study Approach

Yusuf Arayici, Timothy Onyenobiand Charles Egbu (2012). *International Journal of 3-D Information Modeling (pp. 55-73).*

www.irma-international.org/article/building-information-modelling-bim-facilities/62570

Framework for Graphical User Interfaces of Geospatial Early Warning Systems

Martin Hammitzsch (2013). *Geographic Information Systems: Concepts, Methodologies, Tools, and Applications (pp. 449-464).*

www.irma-international.org/chapter/framework-graphical-user-interfaces-geospatial/70455

Mid-Tropospheric Flow Characteristics of Intense Precipitation Events in the Southeastern USA

Walker Skeeterand Jason Senkbeil (2020). *International Journal of Applied Geospatial Research (pp. 10-23).*

www.irma-international.org/article/mid-tropospheric-flow-characteristics-of-intense-precipitation-events-in-the-southeastern-usa/257768