Open Source Social Network Analytics Toolkit

Xinyue Ye, Founding Director, Computational Social Science Lab

Associate Professor, Department of Geography & School of Digital Sciences

Kent State University, Kent, OH 44242

Spatial and social network analyses correspond to the notions of space and place, respectively. Moreover, beyond academia, government agencies, practitioners, as well as engaged citizens are increasingly recognizing the linkages between social space and physical environment. Citizens produce, assemble, and diffuse a large variety of social media information for enhancing situational awareness especially during emergency management and disaster relief. It is important to discover, track, summarize, and even predict popular topics and events occurring in the social network in the space-time context. At the same time, it is very useful that a series of "what if" scenarios can be developed to estimate the meme diffusion. However, spatial social scientists have been slow to adopt and implement new methods for social media data analysis due to the lack of open source software packages, which become a major impediment to the promotion of human dynamics research. The availability and widespread use of source codes will play a critical role in the adoption of new perspectives and ideas enhancing spatial social network analytics. Open Source Social Network Simulator implements the methodological advances in an open source environment of Python for exploring spatial meme diffusion, using twitter data as the case study. Synthesis of spatial and social network analytics has contributed to our understanding of geography of social media such as the spatiotemporal effect in the information diffusion process. More toolkits are needed to interface the open source revolution and human/socioeconomic dynamics analysis seeking cross-fertilization between these two fast-growing communities. The methods are built in open source environments and thus are easily extensible and customizable. The open source movement can also facilitate the explosion of the social media analytics routines by increasingly easier development processes with powerful scripting language environments.

Besides analyzing spatial and temporal characteristics of social media data, some studies focused on mining the actual content of social media messages to improve knowledge about situations. Social media users who are geographically nearby to the events are more possible to provide useful information. Space, time, content, and network are all important attributes of

social media data and should be fully used to gain insights into situational awareness. At the same time, agent-based modeling has been extensively used to predict human behavior like posting, forwarding or replying a message with regard to topics and sentiments. Spatiotemporal visualization can be applied to intuitively reflect the complex process of information diffusion. These visualization techniques allow analysts to iteratively and interactively explore the dataset and thus gain deeper understanding of the origination, propagation, and clustering of information. Our toolkit aims to integrate network generator, network analysis, community detection, and information diffusion into one open source package to conduct computing and simulation on both artificial and real-world networks.

This open source toolkit aims to explore the pattern of information diffusion over time on the social network, so it displays a complete process that starts with preparing a network and ends with demonstrating how information propagates over the network. There are four modules in this toolkit, including Network Generator, Network Analysis, Community Detection, and Information Diffusion. The related functions are described below.

Modules and Description

Modules	Function description
Network Generator	Generate networks with different structures based on parameter settings
Network Analysis	Examine the characteristics of networks or nodes
Community Detection	Detect the distinct groups in a network
Information Diffusion	Provides various algorithms to select seed nodes and opinion leaders to demonstrate information diffusion over network in the independent cascade model and the linear threshold model

This open-source package can promote collaboration among researchers who want to improve current functions or add extensions to address specific research questions on spatial meme diffusion. Based on the strength of interactive visualization techniques, this research stresses the need to study the geography of social media. On this basis, the sincere hope here is that this dialogue between social network and GIS will embrace the real world challenges and opportunities of big social data.