Chapter 11

Exploiting Enriched Knowledge of Web Network Structures: Chaining Maltego Tungsten, NCapture and NVivo, and NodeXL

Shalin Hai-Jew
Kansas State University, USA

ABSTRACT

Understanding Web network structures may offer insights on various organizations and individuals. These structures are often latent and invisible without special software tools; the interrelationships between various websites may not be apparent with a surface perusal of the publicly accessible Web pages. Three publicly available tools may be “chained” (combined in sequence) in a data extraction sequence to enable visualization of various aspects of http network structures in an enriched way (with more detailed insights about the composition of such networks, given their heterogeneous and multimodal contents). Maltego Tungsten™, a penetration-testing tool, enables the mapping of Web networks, which are enriched with a variety of information: the technological understructure and tools used to build the network, some linked individuals (digital profiles), some linked documents, linked images, related emails, some related geographical data, and even the in-degree of the various nodes. NCapture with NVivo enables the extraction of public social media platform data and some basic analysis of these captures. The Network Overview, Discovery, and Exploration for Excel (NodeXL) tool enables the extraction of social media platform data and various evocative data visualizations and analyses. With the size of the Web growing exponentially and new domains (like .ventures, .guru, .education, .company, and others), the ability to map widely will offer a broad competitive advantage to those who would exploit this approach to enhance knowledge.

DOI: 10.4018/978-1-4666-6493-7.ch011
INTRODUCTION

Since the advent of the World Wide Web (WWW) in 1991 and its popularization, researchers have been using this platform to glean publicly available information. Advancements in software tools and research methods, such as network analysis, have enabled broader publics to exploit so-called open-source (publicly available) intelligence (OSINT). The practices of network analysis borrow from math, computer science, and sociology, to depict interactive structures and power relationships in social systems based on interrelationships.

These relationships are often depicted as node-link (vertex-edge) structures, in 2D or 3D. A node (vertex) represents an entity; a link (edge or line) represents some type of relationship. If there are no line ends, the graph is an undirected one; if there are arrows on the line ends, the graph is a directed one (with an indication of the direction of the relationship and the presence or absence of reciprocity). The graph visualizations from network analysis are drawn using a dozen different layout algorithms, all with differing looks-and-feel and levels of analysis. There are some simple dynamics that apply to some classic depictions, such as the core-periphery dynamic: those in the core are often the nodes with high power influence (high in-degree and low out-degree) while those in the periphery are not as influential in that particular social network. Nodes which are pendants (those connected by only one tie) or isolates (those connected by no ties) are considered not very influential. Within networks, there may be cliques (or subclusters or islands) which are more densely connected to each other and then more sparsely connected to the rest of the network. Bridging nodes, those which connect often disparate communities, may have outsized power even if they have few ties—because of their role in connected networks which would not be in communications otherwise. There are other ways that networks are analyzed. There are a range of possible ties that are defined dyadically, quadratically, and so on, that may indicate differential power. Some networks are understood as pair-wise dyadic interactions between two nodes and then the interactions between these pairs linked into various networks; the decisions made by each node results in self-organizing or emergent (some suggest “rhizomic”) behavior that may be seen as patterns at the larger social network levels. In other words, larger macro patterns of cascading behavior may be seen. One example is the phenomena of a meme “going viral” or a product or service gaining in popularity from electronic word-of-mouth percolation.

For example, a node which mediates between two other nodes which do not communicate directly has an oversized sense of power by being the go-between (in a phenomena known as “tertius gaudens” or “the third who benefits” or “the third who rejoices”). Motif censuses may be conducted on various networks to understand both global and node-level insights about that network, from a structural perspective. At a global level, diversity may be seen as promoting resilience but also potentially increased levels of strife. A monolithic network is seen as less robust and less adaptable but also potentially with lowered levels of strife (given the assumption of homophily or people clustering around others like themselves). A linear chain of nodes is seen to be fragile because the removal of any one node can mean a disconnection between numbers of its members. A “star” or a “wheel” (hub and spokes) is a network with an individual at the center who connects all the others; in this context, the power resides in the node at the center (the hub, not the spokes). In a mesh connection, such networks can be more resilient and robust because its members can work around the removal of a wide number of its nodes; there is not a single point-of-failure. A sparse network is thought to be less resilient than a dense (deeply inter-connected) one. This is not to say that structure is destiny. Rather, the theory is that there are interaction effects between structure and behavior, with each affecting the other.