Analyzing a Social Networking Website

Nielsen, J. (2011)
Advisor: Dr. Jessica Libertini

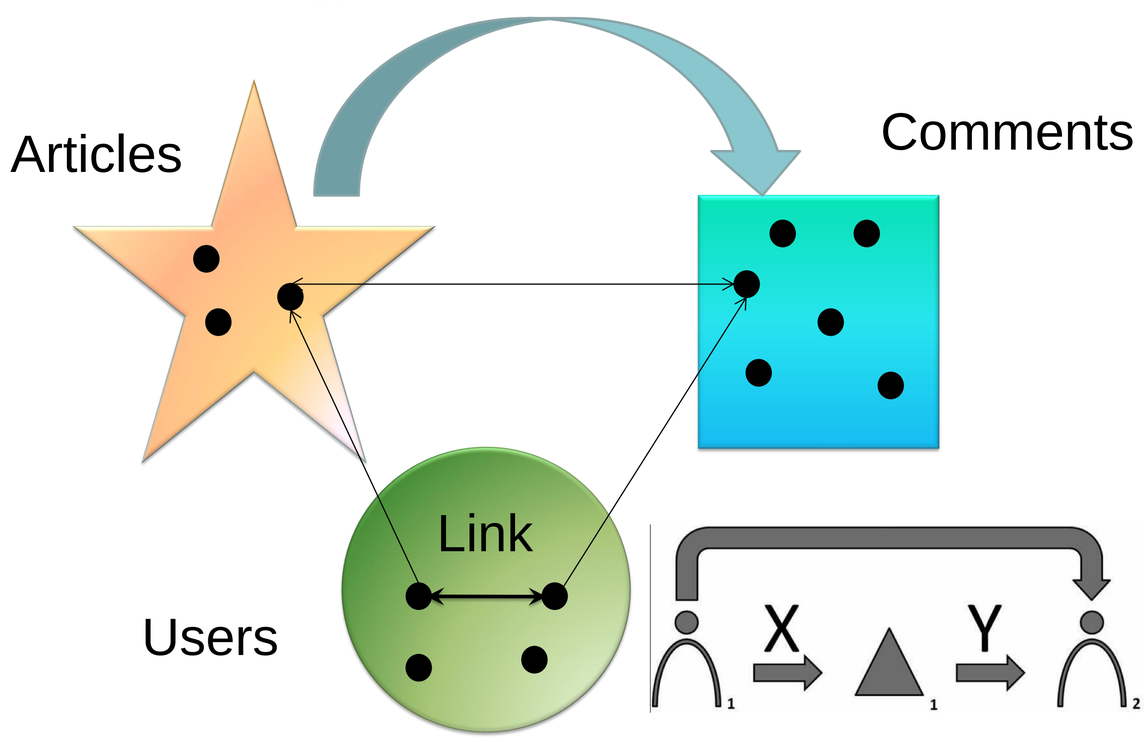
Abstract

For the Cumberland Valley Math Modeling Competition at Shippensburg University in September 2011, my three-person team was given 23 hours to model a 2-week data set from the social networking website Digg.com in order to locate the website's most influential user. After defining the most influential user to be the Agent node that ranked highest in centrality measures such as Hub or Betweenness, we found that a user by the name of "MrBabyMan" had been the most influential user during the 2-week data period. A binary matrix can represent links between any types of node class, thus this process can be used to find influential nodes in collaboration, transportation, insurgent, or any other network. This allows for more effective targeting of those central nodes in order to disrupt or impact the network and is especially applicable to military operations.

Background

My team was given data from Digg.com including its users, posted articles, comments, and likes during a 2-week period and we were tasked to find the website's most influential user without using likes as the primary metric. Our group then had to present its solution in a 2-page executive summary while defending the usefulness of our model.

Methodology



Our group immediately recognized social network analysis as one of the best approaches to finding influential users on Digg.com. Although the website data was composed of 4 unique types– articles, comments, likes, and users– we decided that they could be transformed into a purely social network whose highest centrality nodes could be interpreted as the website's most influential users. Now armed with the task of building a user by user network from the 4 different node types and the purpose of improving advertisement targeting within the network, our group moved on to convert the given data into something usable by *ORA.

*ORA is a program commissioned by the Department of Defense that can construct networks, conduct matrix operations on them, and generate useful visualizations. However, the data from Digg.com was given to us as a tabulated .xls document and needed to be transformed into a binary matrix .csv file in order to be imported with *ORA. This was the most challenging portion of the competition because it required 14 hours to create a binary matrix for each of the 4 node classes. Once this had been accomplished, we used the principles of network science to multiply each user by class matrix by its transpose and treat their sum as the entire social network of Digg.com

Results

One of the most commonly used centrality measures is Betweenness because it describes how often a node occurs within the geodesic paths between two nodes that do not share a direct link with each other. However, we encountered a network density problem in which there were few links between users relative to the total number of possible links. In order to overcome this challenge, we decided to take the most influential user as the one that consistently ranked highest in Degree, Hub, and Eigenvector centrality: "MrBabyMan" was the most central node in all 3 of these measures, as well as in a few others.

Application

Using only data and matrix algebra software, it is possible to automate this process and apply it to any number of different networks. Transportation, area defense, insurgency, information, pollution, and political networks can all be quickly analyzed for key terrain, leaders, or ideas. This can be particularly useful in contemporary Army missions that focus on counter-insurgency and full spectrum operations rather than kinetic warfare.

Insurgency networks, such as al Qaeda in Afghanistan or the Malaysian National Liberation Army, are very often decentralized in order to maximize their battle space while minimizing the importance of any one location. Using traditional strategy is very ineffective against these insurgencies. However, central nodes identified through intelligence and network analysis can be efficiently and effectively exploited by the host nation. For example, the Briggs Plan in Malaysia employed siloing by reducing the number of food nodes that local insurgents could access, while focusing security measures on those few links that remained.

Network analysis offers cost-effective alternatives that exploit an insurgency's leadership, security, or logistics by eliminating key nodes that link otherwise isolated sections of the insurgency network. By employing network science and analysis, the US Army can increase its operational capability while reducing overall costs.

References.
Carley, Kathleen M. "*ORA." Carnegie Mellon University, 2011.
LTC McCullough and MAJ Anthony Johnson. "Advanced Network Analysis and Targeting Course." Wiley Preprint, 2011.
Photos by U.S. Army, BBC Hulton Picture Library