GLOBAL RESEARCH IMMERSION PROGRAM
FOR YOUNG SCIENTISTS

Community detection in Internet Structure

autonomous system

Xiangyue Que
University of Illinois Urbana-Champaign
anneqxy2003@gmail.com

Abstract
Social network analysis (SNA) is a research method that uses networks and graph theory to study social structures and the relationships between people or organizations. The goal is to understand a community by mapping the relationships between its members and identifying key individuals, groups, and associations. SNA can be used to visualize many types of social structures, including social media networks, business networks, and friendship networks.

Introduction

Social network analysis (SNA) is the process of investigating social structures through the use of networks and graph theory. It characterizes networked structures in terms of nodes (individual actors, people, or things within the network) and the ties, edges, or links (relationships or interactions) that connect them. Examples of social structures commonly visualized through social network analysis include social media networks, meme spread, information circulation, friendship and acquaintance networks, peer learner networks, business networks, knowledge networks, difficult working relationships, collaboration graphs, kinship, disease transmission, and sexual relationships. These networks are often visualized through sociograms in which nodes are represented as points and ties are represented as lines. These visualizations provide a means of qualitatively assessing networks by varying the visual representation of their nodes and edges to reflect attributes of interest.

Social network analysis has its theoretical roots in the work of early sociologists such as Georg Simmel and Émile Durkheim, who wrote about the importance of studying patterns of relationships that connect social actors. Social scientists have used the concept of ”social networks” since early in the 20th century to connote complex sets of relationships between members of social systems at all scales, from interpersonal to international.

In the 1930s Jacob Moreno and Helen Jennings introduced basic analytical methods.[23] In 1954, John Arundel Barnes started using the term systematically to denote patterns of ties, encompassing concepts traditionally used by the public and those used by social scientists: bounded groups (e.g., tribes, families) and social categories (e.g., gender, ethnicity).

Starting in the 1970s, scholars such as Ronald Burt, Kathleen Carley, Mark Granovetter, David Krackhardt, Edward Laumann, Anatol Rapoport, Barry Wellman, Douglas R. White, and Harrison White expanded the use of systematic social network analysis.

Beginning in the late 1990s, social network analysis experienced a further resurgence with work by sociologists, political scientists, economists, computer scientists, and physicists such as Duncan J. Watts, Albert-László Barabási, Peter Bearman, Nicholas A. Christakis, James H. Fowler, Mark Newman, Matthew Jackson, Jon Kleinberg, and others, developing and applying new models and methods, prompted in part by the emergence of new data available about online social networks as well as ”digital traces” regarding face-to-face networks.

Background

The Internet is a network of networks*, and autonomous systems are the big networks that make up the Internet. More specifically, an autonomous system (AS) is a large network or group of networks that has a unified routing policy. Every computer or device that connects to the Internet is connected to an AS.

Imagine an AS as being like a town’s post office. Mail goes from post office to post office until it reaches the right town, and that town’s post office will then deliver the mail within that town. Similarly, data packets cross the Internet by hopping from AS to AS until they reach the AS that contains their destination Internet Protocol (IP) address. Routers within that AS send the packet to the IP address. Every AS controls a specific set of IP addresses, just as every town’s post office is responsible for delivering mail to all the addresses within that town. The range of IP addresses that a given AS has control over is called their ”IP address space.” Most ASes connect to several other ASes. If an AS connects to only one other AS and shares the same routing policy, it may instead be considered a subnetwork of the first AS. Typically, each AS is operated by a single large organization, such as an Internet service provider (ISP), a large enterprise technology company, a university, or a government agency.

An AS routing policy is a list of the IP address space that the AS controls, plus a list of the other ASes to which it connects. This information is necessary for routing packets to the correct networks. ASes announce this information to the Internet using the Border Gateway Protocol (BGP). .

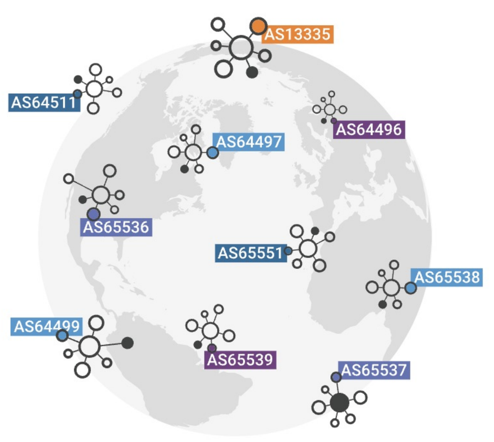
A specified group or range of IP addresses is called ”IP address space.” Each AS controls a certain amount of IP address space. (A group of IP addresses can also be called an IP address ”block”.)

Imagine if all the phone numbers in the world were listed in order, and each telephone company was assigned a range: Phone Co. A controlled numbers 000-0000 through 599-9999, and Phone Co. B controlled numbers 600-0000 through 999-9999. If Alice calls Michelle at 555-2424, her call will be routed to Michelle via Phone Co. A. If she calls Jenny at 867-5309, her call will be routed to Jenny by Phone Co. B.

This is sort of how IP address space works. Suppose Acme Co. operates an AS and controls an IP address range that includes the address 192.0.2.253. If a computer sends a packet to 192.0.2.253, the packet will eventually reach the AS controlled by Acme Co. If that first computer is also sending packets to 198.51.100.255, the packets go to a different AS (although they may pass through Acme Co.’s AS on the way).

When networking engineers communicate which IP addresses are controlled by which ASes, they do so by talking about the IP address ”prefixes” owned by each AS. An IP address prefix is a range of IP addresses. Because of the way IP addresses are written, IP address prefixes are expressed in this fashion: 192.0.2.0/24. This represents IP addresses 192.0.2.0 through 192.0.2.255, not 192.0.2.0 through 192.0.2.24.

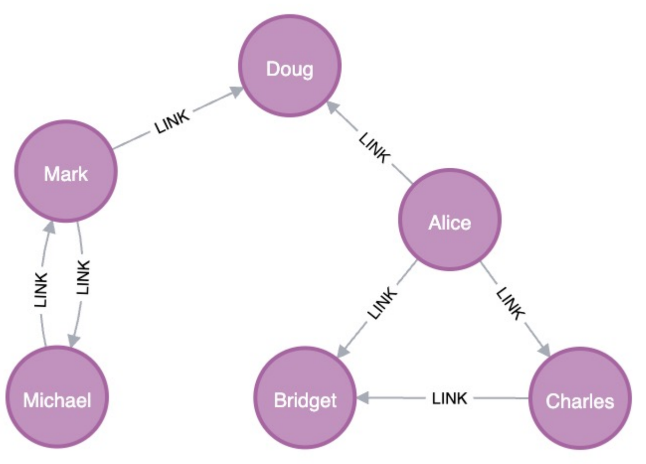
ASes announce their routing policy to other ASes and routers via the Border Gateway Protocol (BGP). BGP is the protocol for routing data packets between ASes. Without this routing information, operating the Internet on a large scale would quickly become impractical: data packets would get lost or take too long to reach their destinations.



Methods

Louvain Algorithm: Louvain is an unsupervised algorithm (does not require the input of the number of communities nor their sizes before execution) divided in 2 phases: Modularity Optimization and Community Aggregation. After the first step is completed, the second follows. Both will be executed until there are no more changes in the network and maximum modularity is achieved. Louvain will randomly order all nodes in the network in Modularity Optimization. Then, one by one, it will remove and insert each node in a different community C until no significant increase in modularity (input parameter) is verified.

Modularity metric is a metric that allows you to evaluate the quality of a community detection. Relationships of nodes in a community C connect to nodes either within C or outside C. Graphs with high modularity have dense connections between the nodes within communities but sparse connections between nodes in different communities.



Results

To effectively demonstrate the evolution of community structures over a span of 25 years, we will utilize advanced detection algorithms and sophisticated visual tools. These technologies will allow us to accurately identify and analyze changes within the communities, providing detailed insights into their development and transformation over time. By leveraging these methods, we can create comprehensive visualizations that clearly illustrate the dynamic nature of community structures, highlighting key trends and significant shifts that have occurred throughout the 25-year period.

Conclusions

In conclusion, by employing advanced detection algorithms and sophisticated visual tools, we can effectively demonstrate the evolution of community structures over a span of 25 years. These technologies enable us to accurately identify and analyze changes within communities, providing detailed insights into their development and transformation over time. Leveraging these methods allows us to create comprehensive visualizations that clearly illustrate the dynamic nature of community structures, highlighting key trends and significant shifts that have occurred throughout the 25-year period.