ABSTRACT

A rise in criminal activity on the dark web, including the sale of personal data, may be attributed to a deterioration in state security and privacy protection. In addition to security issues, researchers looking into network security need access to the scant information currently available about the dark web. Based on these problems, data on the dark web can be further investigated through content analysis. Finding information about content on the dark web is advised to be done through content analysis. Content analysis (URL) is used to determine the centrality, density, and modularity values of each dark web node.

The centrality value is determined using three approaches: degree, relatedness, and closeness. These approaches yield distinct dominant contents. Examining the values of the top 20 URLs allows one to confirm that, according to the betweenness value calculation, URL 397 is the primary URL affecting other nodes' relationships. Furthermore, it can be inferred from the closeness value calculation that URL 397 is the node with the highest closeness value to other nodes. But since every node has the same closeness value, there isn't a dominant URL node. The density is quite dense with relationships between nodes, a lot of data distribution, and a strong community structure, according to further analysis of the density (value 0.145) and modularity (value 0.674). The relationship between density and modularity values is inversely proportional: a network with low density has high modularity because it has a high degree of data distribution, as evidenced by the dark web content analysis value.

Keywords: *Analytics, Crawling, Dark Web, TOR, Centrality, Density, Modularity.*