Unsupervised Graph Network Discrimination

In recent years, network analysis techniques have evolved(Chen et al., 2019; Williamson & Tec, 2019) in proportion to the rapid growth of real world networks. Much research has been done on networks such as social networks (Bródka et al., 2011; Culotta & Cutler, 2016), communication networks (Mccallum, 2007), citation networks etc. As time progresses, these networks become bigger and more complex, consequently holding vast amounts of interesting information which can be used for various purposes. Network analysis techniques include, but are not limited to, using random graph models to capture or derive the properties of real world networks (Williamson & Tec, 2019), subgraph isomorphism (Cordella et al., 2004), graph simulation etc. A significant portion of the research conducted on networks has been on centralities (Newman, 2010) to deduce, “which” are the more important nodes in the graph i.e.. central nodes or nodes that propagate most information, and how traffic flows through them. In turn, calculated centralities of various network graphs have been used in many other research fields as well. For example, analyzing social networks, such as tweet classification (Hussain & Islam, 2016), detecting political discussion practices (Miller et al., 2015) and many other such applications (Cohn et al., 2019; Rossman et al., 2010; Yang & Liu, 2008); has involved the generation of multiple centralities. Centralities have also been used in analyzing road traffic networks, such as finding road network patterns (Zhang et al., 2011), tourism management (Lee et al., 2013) etc.. Another important application of centrality measurements is in analyzing biological networks, as can be seen in (Bell et al., 1999; Joyce et al., 2010; Narayanan, 2005; Park & Kim, 2009).

Centralities such as Degree, Closeness, Betweenness, Crossclique, Pagerank, etc. have all been proposed and developed (Crucitti et al., 2006) over the years to answer “Which node is the most influential?” in various different applications (Landherr et al., 2010). “Significance” or “importance” of nodes varies from one context to another. For example, in some cases, it is essential to identify which node(s) propagate more information locally or globally in a graph(Newman, 2010), whereas in other cases, detecting the central node(s) might be of more value (Chiu et al., 2016). Centralities can, therefore, be seen as features/characteristics of a graph; thus, it is possible to discriminate between networks based on their centrality values (Wang & Krim, 2012) //*citation added manually.*

In this paper, our main concern is “What information can we extract from a given, unknown graph of a network?” Calculating various centralities from a graph is possible, but this raises the question of what can be concluded from this information? Is it possible to discriminate between types of networks based on only the centrality values of their nodes? Or, Can unsupervised learning techniques applied to unknown graphs produce meaningful context about the network in question? Being able to find meaningful insights about an unidentified network has many practical uses, such as deducing, whether it is a criminal network or not, or if the network was part of a successful political campaign, or whether the network is an ego network of a socially influential person.