Proposal for Studying Monophily in Real Social Networks

Rain Huang

Altenburger and Ugander (2018) introduced the concept of “monophily” in their paper to represent the phenomenon of individuals with extreme preferences for a particular attribute that is possibly unrelated with their own attribute. As a result, friends-of-friends are more likely to be similar. While “homophily” focuses on a bias in attribute preferences for similar others, “monophily” considers the excess variance/overdispersion in attribute preferences. They define this excess variance as observing more variance than expected under a statistical model of interaction preferences with homophily as a bias parameter. To examine the impact of monophily on network structure, they developed an extension of the stochastic block model (SBM), which is a classic model of biased preferences in networks, by introducing an additional parameter to quantify overdispersion. When , there is no excess variation, meaning nodes in class r have similar proportions of in-class and out-class neighbors. represents excess variance beyond the conventional model. The modified model (oSBM) can capture homophily and monophily separately and by utilizing it in node/relational inference methods, prediction performance in the case of overdispersion of attribute preferences can be greatly improved with weak or even no homophily. They investigated 4 different empirical social networks where monophily might be present – gender in FB100 networks (Traud et al., 2012) and Add Health networks (Resnick et al., 1997), which may exhibit weak homophily, and political affiliations of online blogs (Adamic & Glance, 2005) and contact network of terrorist group members and non-members in the Noordin Top Terrorist Network (Roberts & Everton, 2011), which are known to be highly homophilous. They applied 4 node classification methods to predict node attribute in each of the four networks. 2 classification methods, the one-hop MV classifier and the ZGL method, are based on a node’s one-hop (immediate) relations, while the other 2, two-hop MV classifier and LINK-logistic regression, are based on two-hop (neighbor of neighbor) relations. The two-hop methods (two-hop MV and LINK) performed consistently better than one-hop ones in predicting gender on Facebook networks, and as well as the other methods in predicting political affiliation. I would like to investigate more empirical social networks characterized by various structural patterns in attribute (differing levels of homophily and monophily), apply the oSBM framework to them to find h (homophily index) and (monophily) estimates, and perform node classification using the 4 classification methods. I want to compare the performance of the one-hop and two-hop methods when used on networks with different levels of homophily and monophily.