

1135-68-2777

Daniel B Larremore* (daniel.larremore@colorado.edu), **Leto Peel** and **Aaron Clauset**.

The ground truth about metadata and community detection in networks.

It is common to evaluate the performance of community detection algorithms by their ability to find so-called ground truth communities. This works well in synthetic networks with planted communities because these networks' links are formed explicitly based on those known communities. However, there are no planted communities in real-world networks. Instead, it is standard practice to treat some observed discrete-valued node attributes, or metadata, as ground truth. We show that metadata are not the same as ground truth and that treating them as such induces severe theoretical and practical problems. We prove that no algorithm can uniquely solve community detection, and we prove a general No Free Lunch theorem for community detection, which implies that there can be no algorithm that is optimal for all possible community detection tasks. However, community detection remains a powerful tool and node metadata still have value, so a careful exploration of their relationship with network structure can yield insights of genuine worth. We illustrate this point by introducing two statistical techniques that can quantify the relationship between metadata and community structure for a broad class of models. (Received September 26, 2017)