

1 Conclusion

In this study, we have demonstrated that the communities observed in online social networks are highly question-dependent, and that different definitions of communities reveal different relationships between users. More importantly, we have shown that these different views of the network are not revealed by using the structural network alone. We found that community structure differs across community types on both the macro (e.g. number of communities and their size distribution) and micro (e.g. specific membership) scale. We demonstrated that boundaries between communities do represent meaningful internal/external divisions, with conversations and topics tending to be most highly concentrated within communities than without. We found this to be the case even when the community- and edge-types did not agree.

Our findings have important implications to a common problem in social network analysis: identification of influential individuals. Many network measures of influence are based on the various types of centrality (degree, betweenness, closeness, eigenvector, etc.) [?]. Most centralities explicitly depend on the structure of the network under consideration. But we have seen in our study that a structural network is not sufficient to capture user interaction in online social media. Thus, a naive application of centrality measures to a structural network for influence detection may give rise to erroneous results. This result has been explored previously [?], and our work further highlights its importance. Moreover, weighted generalizations of these centralities might lead to better insights about who is actually influential in a social network.

Our work has introduced a novel use of transfer entropy for the detection of activity-based communities. In previous work on online social media [?], transfer entropy was found useful for identifying influence between users. It is thus interesting that we find influence tends to be higher across community boundaries than within them. This counterintuitive result warrants further investigation. In addition, more rigorous choices of both the lag and coarsening resolution based on methods for model selection should be explored [?].