A Study of Explainable Community Level Features

Sophie Sadler and Daniel Archambault
Swansea University, UK

Derek Greene
School of Computer Science, University College Dublin, Ireland

Abstract

Community finding algorithms are complex, stochastic algorithms used to find highly connected groups of individuals in a graph. As with "black-box" machine learning approaches, they provide little explanation or insight into their outputs. Inspired by work in explainable artificial intelligence (XAI), we look to develop post-hoc explanations for community finding algorithms. Specifically, we aim to identify features that indicate whether a set of nodes comprises a community or not. We evaluate our model-agnostic methodology, which selects interpretable features from a longlist of candidates, on three well-known community finding algorithms.

Introduction

Community finding is an important task for gaining insight into the network structure. Networks are normally used to represent relational, non-Euclidean data. Data points, known as nodes, are connected by edges which represent relationships between the nodes. Communities are loosely defined as sets of nodes with high connectivity within the community and sparser connections to nodes outside of the community.

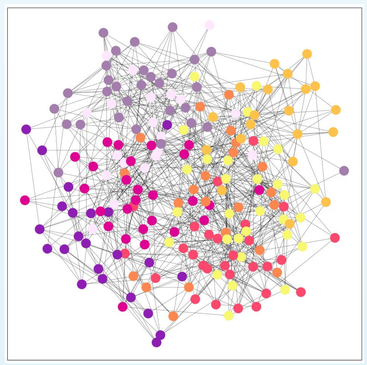

Figure 1: A network with communities shown with colour.

Existing community finding algorithms provide little insight beyond the identification of the communities themselves. This, along with their stochasticity, leave it uncertain as to why the algorithm has identified a certain set of communities.

In the wider field of machine learning, explainability has become imperative to avoiding hidden biases or incorrect assumptions in "black-box" approaches. Where it has proved hard to develop a "transparent" model (i.e., a model where the inner workings are easily understood), "post-hoc" explanations have been developed as an alternative. This approach involves generating explanations for outputs after the model has already been trained and applied on the data.

One such "post-hoc" method is to identify interpretable features which a domain expert can easily recognise and understand, which is the approach we take in our work.

Methodology

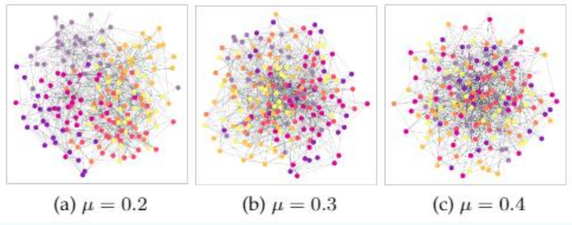

Figure 2: Example graphs with 200 nodes at the three μ (mixing parameter) values used. Increasing the mixing parameter increases the prevalence of edges between communities.

We perform a set of experiments on three algorithms at different network μ values. A large set of synthetic graphs on which to perform the experiments are generated using the LFR benchmark method, for which μ is the mixing parameter which determines the separation of communities. Then for each experiment, we run the chosen algorithm on the set of graphs with a specific μ value.

For each graph, we identify the set of unique communities found across many runs of the algorithm. One node may appear in many different communities within this set, as the community structure may have been identified differently across different runs. However, each community will appear only once in the dataset.

We then compute the value of our longlist of features for these communities. This gives us our set of features for the "real community" labelled datapoints. A rewiring process is used to adjust the network structure and the features are recalculated, giving us our set of features for the "fake community" labelled datapoints. The structure of the community has changed in the rewiring, resulting in new values for the features.

Finally, we train a random forest classifier to distinguish between the "real" and "fake" community classes using the features we have precalculated, and analyse the permutation importances of respective features. To do this, t-tests with Bonferroni-Holm corrections are used to compare feature importances and identify pairs with significant difference in their level of importance.

The longlist of features chosen for our experiments are as follows: relative density, relative diameter, relative pathlength, relative degree, relative betweenness, relative closeness, cut ratio and internal-external.

Acknowledgements

This work is supported by the UKRI AIMLAC CDT, funded by grant EP/S023992/1 and by Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289P2.

Experiments

We perform our experiments on the following stochastic algorithms, designed for detecting partitions (i.e., non-overlapping communities): Louvain [1], Infomap [8], Label Propagation (LPA) [6]. We use 3 μ values: 0.2, 0.3, 0.4. However, we omit LPA on μ = 0.4 since it often classifies all nodes in the graph as belonging to a single community.

Distributions of permutation importance for the features on each experiment are displayed below in figure 3. These distributions are constructed using permutation importance values calculated for each graph in the dataset. From our experimental results, we see qualitatively that the cut ratio and internal-external metrics are consistently the most important features in distinguishing the "real communities" from the "fake communities". However, as the μ value increases, there is evidence to suggest that the relative betweenness may also have some importance.

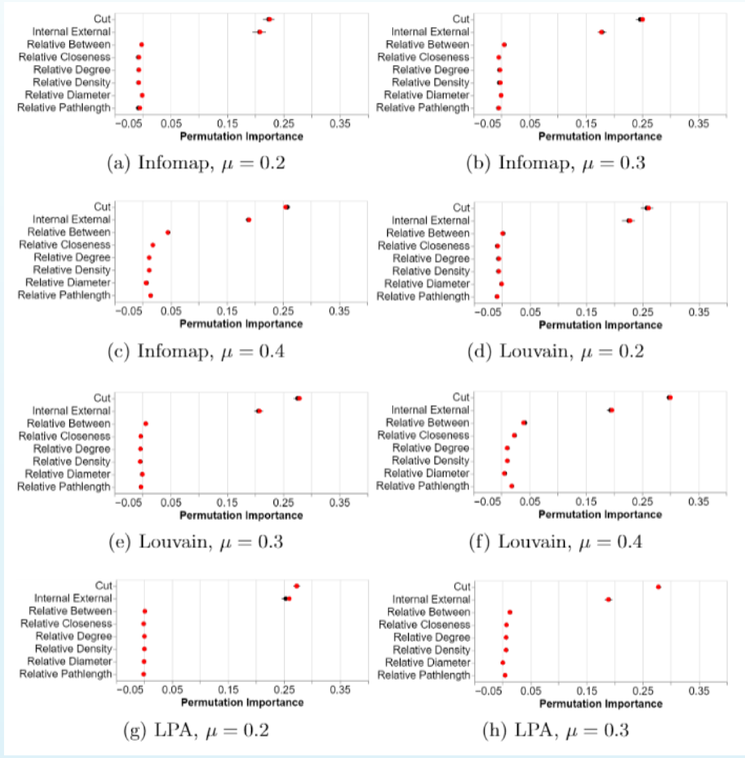

Figure 3: Results of the community feature experiments. Plots are of permutation importance of the metrics. Mean indicated as a black dot and median as a red dot. Lines indicate 95% bootstrapped confidence intervals.

Conclusion

We have presented the methodology and results of an experiment to determine features that can be used to explain detected community structure in networks. Our experiment tested features which were calculated on sets of nodes that could form a possible community. We find that cut ratio and the internal-external ratio are the most informative features. As the mixing of the communities increases, relative betweenness increases in importance. Moving forward, we would like to perform the necessary HCI and visualisation work to construct explainable network analysis systems that help real users with their tasks.

References

[1] Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of com-munities in large networks. Journal of statistical mechanics: theory and experiment2008(10), P10008 (2008)

[2] Fortunato, S.: Community detection in graphs. Physics reports 486(3-5), 75–174(2010)

[3] Lancichinetti, A., Fortunato, S.: Community detection algorithms: A comparative analysis. Physical Review E - Statistical, Nonlinear, and Soft Matter Physics 80(5),1–12 (2009)

[4] Lancichinetti, A., Fortunato, S.: Consensus clustering in complex networks. Scientific reports 2(1), 1–7 (2012)

[5] Lancichinetti, A., Fortunato, S., Radicchi, F.: Benchmark graphs for testing community detection algorithms. Physical Review E - Statistical, Nonlinear, and Soft Matter Physics78(4), 1–6 (2008)

[6] Raghavan, U.N., Albert, R., Kumara, S.: Near linear time algorithm to detect community structures in large-scale networks. Physical review E76(3), 036106(2007)

[7] Ribeiro, M.T., Singh, S., Guestrin, C.: "Why should I trust you?" Explaining the predictions of any classifier. Proc. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining pp. 1135–1144 (2016)

[8] Rosvall, M., Bergstrom, C.T.: Maps of random walks on complex networks reveal community structure. Proc. National Academy of Sciences of the United States of America 105(4), 1118–1123 (2008)

[9] Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature393, 440–442 (1998)