Template for preparing your research report submission to PNAS using Overleaf

Jan Kenda[a] **and Andraž De Luisa**[b]

[a]University of Ljubljana, Faculty of Computer and Information Science, Ljubljana, Slovenia, jk—-@student.uni-lj.si ; [b]University of Ljubljana, Faculty of Computer and Information Science, Ljubljana, Slovenia, ad9366@student.uni-lj.si

This manuscript was compiled on May 11, 2020

Community structures are quite common in real networks, especially in social networks. We are interested in the robustness of such groupings of nodes into communities and in the importance of single nodes in their respective communities. We analyse how likely are these communities to break apart if we remove some nodes from them, and how to find such nodes that are most important for the survival of their community. We evaluate the performance of the proposed methods on some benchmark graphs and the GitHub social network.

communities | robustness | node centrality |

Sharing ideas, problems and solutions is a fundamental reason why humans are as advanced as they are today. This is true in all fields of science: biology, physics, mathematics and potentially even more so in computer science. With the advancements in technology and specifically the internet in the last couple of decades, cooperation between researchers and innovators has never been easier. This is in large part due to Git and other version control systems, which allow both researchers and software developers to interact and collaborate on projects with ease, in turn leading to faster learning, innovation and development of scientific and technological advancements.

But this network has a big drawback: it is a social network. And as a social network, it has its linchpins - key members of a society which hold everything together. Without them the network might fall apart. These are the people who introduce others. The keys that connect people into a community. And this is what we are interested in.

If we are able to find a community structure in the GitHub network of collaborators (1), we can then try to measure how "well connected" each community is. Using this measure as a criterion, we can then start looking for linchpins - in this case nodes, which if removed, would destabilize the community structure of the network as much as possible. In the real world social network this means removing individuals that we think contribute most to the development of new ideas and projects. Not necessarily provide solutions themselves, but by connecting others and creating (or just being the core of) communities, contribute to further technological advancements the most.

In this paper we analyse the ability of several node centrality measures to highlight these crucial nodes for the robustness of the community structure in social networks.

Related work

Node centrality: - bridgeness (2) - betweenness (3) - pagerank (4) - community centrality (5) - shapley values centrality (6), (7)

Community robustness: - conductance (8) - pairwise similarity (9) - variation of information in perturbed networks (10)

Benchmark networks: - Girvan-Newman (11) - Lancichinetti (LFR) (12) - Affiliation graph model (13)

Methods

- find communities in network (if we don't know them) - rank nodes with respect to a selected measure - remove best nodes - compare new communities with previous

Results

Discussion

Conclusion

1. Rozemberczki B, Allen C, Sarkar R (2019) Multi-scale attributed node embedding.
2. Jensen P, et al. (2015) Detecting global bridges in networks. *Journal of Complex Networks* 4(3):319–329.
3. Freeman LC (1977) A set of measures of centrality based on betweenness. *Sociometry* 40(1):35–41.
4. Page L, Brin S, Motwani R, Winograd T (1999) The pagerank citation ranking: Bringing order to the web., (Stanford InfoLab), Technical Report 1999-66. Previous number = SIDL-WP-1999-0120.
5. Newman MEJ (2006) Finding community structure in networks using the eigenvectors of matrices. *Physical Review E* 74(3).
6. Michalak T, Aadithya K, Szczepański P, Ravindran B, Jennings N (2014) Efficient computation of the shapley value for game-theoretic network centrality. *Journal of Artificial Intelligence Research* 46.
7. Narahari Y (2011) A shapley value-based approach to discover influential nodes in social networks. *Automation Science and Engineering, IEEE Transactions on* 8:130 – 147.
8. Kannan R, Vempala S, Vetta A (2004) On clusterings: Good, bad and spectral. *J. ACM* 51(3):497–515.
9. II MJB, Katz DM, Zelner JL (2010) On the stability of community detection algorithms on longitudinal citation data. *Procedia - Social and Behavioral Sciences* 4:26 – 37. Applications of Social Network Analysis.
10. Karrer B, Levina E, Newman MEJ (2008) Robustness of community structure in networks. *Physical Review E* 77(4).
11. Girvan M, Newman MEJ (2002) Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* 99(12):7821–7826.
12. Lancichinetti A, Fortunato S, Radicchi F (2008) Benchmark graphs for testing community detection algorithms. *Phys. Rev. E* 78(4):046110.
13. Yang J, Leskovec J (2012) Community-affiliation graph model for overlapping network community detection in *2012 IEEE 12th International Conference on Data Mining*. pp. 1170–1175.

Please provide details of author contributions here.

Please declare any conflict of interest here.

[1]A.O.(Author One) and A.T. (Author Two) contributed equally to this work (remove if not applicable).

[2]To whom correspondence should be addressed. E-mail: author.twoemail.com