WS22-23: Seminar Social Network Analysis

Review report for the paper "k-Centralities: Local Approximations of Global Measures Based on Shortest Paths"

Oleksandr Tarasov
oleksandr.tarasov@stud-mail.uni-wuerzburg.de
Julius-Maximilians-Universität Würzburg

I. Summary of the paper

The paper tackles the problem of efficiently estimating global centrality measures in large graphs by utilizing local information within k-hop neighborhoods. Centrality measures are crucial for analyzing the importance of nodes in graphs and can also be employed to reduce dimensions in multi-dimensional complex networks, thereby simplifying the analysis and enhancing the interpretability of these intricate structures; however, computing traditional centrality measures, such as Betweenness and Closeness, can be computationally expensive, especially for large-scale graphs. Also, results of centrality measures for the current time stamp in real-time networks can lead to irrelevant results. To overcome this challenges, the authors introduce the concept of k-Centralities, where k represents the size of the local neighborhood around a node.

The paper proposes the approach to find k-Betweenness and k-Closeness centralities. These centrality measures are used to approximate their respective global options. The algorithm uses modified BFS version to consider number of shortest paths and their length. The k-Betweenness part calculates the number of shortest paths passing through a node within the k-hop neighborhood, while the k-Closeness part estimates the sum of the shortest path distances to all other nodes within the k-hop neighborhood. The algorithm leverage local information, reducing the computational complexity compared to their global counterparts.

The authors conduct extensive experiments on synthetic and real-world datasets to evaluate the performance of the proposed k-Centrality measures. The results indicate that the k-Centralities provide accurate approximations of the global centrality measures with significant improvements in computational efficiency. But at the same time authors discuss how validity and computation time depend on the network parameters like average path and clustering coefficient. Moreover, the paper discusses the impact of varying the k value on the accuracy of the approximations, providing insights into the trade-offs between computational cost and estimation accuracy.

II. Positive aspects

There are several positive aspects in this paper:

- Profound analysis of the field
- Novel algorithms for local approximation of global centrality measures
- Comprehensive empirical evaluation.

III. Negative aspects

At the same time we can mention what was missed in the paper:

- Selection of the optimal k
- Comparision to existing methods
- Limited variety of real-world datasets

IV. Evaluation of related work

The paper presents the novel approach based on [1] and [2]. The algorithm is mostly taken from [2], but the novelty of the paper are the formulas for the k-betweenness and k-closeness centralities, especially normalization part. The paper discusses related works on traditional global centrality measures and their computational challenges, it falls short in providing a detailed and comprehensive overview of existing local approximation methods and state-of-the-art algorithms for estimating global centrality measures.

V. Detailed evaluation

First of all let‘s consider positive aspects of the paper.

A. Profound analysis

Authors demonstrate an in-depth analysis of the field, showcasing a thorough understanding of the challenges and existing methods in graph analysis, which contributes to the development of the innovative k-Centrality measures presented in the paper. For instance, the author mentioned previous works for calculating centralities like ego networks, pivots and node cuts methods, and community centrality. Also authors mentioned the previous research of the bonded-distance and efficient betweenness centrality algorithms which are exactly the base for k-centrality algorithm.

B. Novelty

The approach for calculating k-Betweenness and k-Closeness centrality measures that were introduced in the paper are innovative centrality measuress that leverage local information within k-hop neighborhoods to approximate traditional global centrality measures. These algorithms offer a fresh perspective on how to balance computational efficiency and accuracy in network analysis, particularly for real-time and large-scale networks.

C. Comprehensive empirical evaluation

The paper presents an extensive set of experiments on both synthetic and real-world datasets to evaluate the performance of the proposed k-Centrality measures. The results not only demonstrate the effectiveness and efficiency of these measures in approximating their global counterparts, but also provide insights into the impact of varying the k value on the trade-offs between computational cost and estimation accuracy. This thorough evaluation lends credibility to the proposed methods and paves the way for future research in graph analysis. Also the authors mentioned the dependency for the time and accuracy from network parameters like average path length. And the last but not list was revealing a network where the algorithm will be totally wrong.

Let us now proceed to address the negative aspects of the topic at hand.

D. Selection of the optimal k

The choice of k, which represents the size of the local neighborhood around a node, is a crucial factor in achieving accurate approximations of global centrality measures. The paper does not provide sufficient guidance or techniques for selecting an appropriate k value based on different graph structures, such as social networks, biological networks, or transportation networks. The lack of a clear methodology for choosing k might limit the practical application of the proposed k-Centrality measures and affect their accuracy in diverse real-world scenarios.

E. Limited variety of real-world datasets

Although the paper includes experiments on both synthetic and real-world datasets, the range of real-world datasets could have been broader. A more diverse set of datasets, such as online social networks, collaboration networks, and transportation networks, would help validate the applicability and generalizability of the proposed k-Centrality measures in different contexts and network types. This would also allow for a better understanding of the performance of the algorithms under varying network characteristics, topologies, and sizes, providing further insights into their effectiveness across a wide range of applications.

F. Comparision to existing methods

While the paper presents a novel approach to approximating global centrality measures, it falls short in providing a comprehensive comparison with other existing local approximation methods in the literature as [3], [4] or [5]. A thorough comparison would not only help highlight the strengths and weaknesses of the proposed k-Centrality measures but also place the work in a broader context within the field, allowing for a better understanding of its contributions to the existing body of knowledge.

VI. Reproducibility

The paper provides sufficient detail and data for reproducing the results. The authors clearly describe the algorithms, datasets, and experimental settings. At the same time, it would be helpful if the authors made the code and datasets publicly available in one repository to facilitate replication and further research.

VII. Editional remarks

To identify potential grammatical errors, I utilized LanguageTool [6], a "multilingual grammar, style, and spell checker" that offers a document checker feature compatible with .docx files. To facilitate this process, I converted the PDF to a .docx format. Upon examination, the tool confirmed that the paper is well-prepared from an editorial standpoint.

VIII. Standard of writing

The text exhibits a high standard of writing. The author effectively explains the core concepts in the introduction, and their writing style, along with the choice of words, remains intuitive and easily comprehensible throughout the subsequent chapters.

IX. Overall judgment

I recommend a weak accept (1) for this paper. The proposed method is innovative and shows promising results, but the paper could benefit from addressing the points mentioned in the "Detailed Evaluation" section.

X. Statement of confidence

I possess a medium level of confidence in my review. While I am familiar with the subject matter and my assessments appear reasonable, I can not assert with complete certainty the novelty of the formulas of the k-betweenness and k-closeness centrality measures, especially with regards to the normalization parts.

References

[1] S. P. Borgatti and M. G. Everett, "A graph-theoretic perspective on centrality," *Social Networks*, vol. 28, no. 4, pp. 466–484, 2006, ISSN: 0378-8733. DOI: http://dx.doi.org/10.1016/j.socnet.2005.11.005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378873305000833.

[2] U. Brandes, "On variants of shortest-path betweenness centrality and their generic computation," *Social Networks*, vol. 30, no. 2, pp. 136–145, 2008, ISSN: 0378-8733. DOI: http://dx.doi.org/10.1016/j.socnet.2007.11.001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378873307000731.

[3] D. Bader and K. Madduri, "Parallel algorithms for evaluating centrality indices in real-world networks," in *Proc. The 35th International Conference on Parallel Processing (ICPP)*, 2006.

[4] D. Eppstein and J. Wang, "Fast approximation of centrality," *Symposium on Discrete Algorithms: Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms*, vol. 7, no. 09, pp. 228–229, 2001.

[5] R. Geisberger, P. Sanders, and D. Schultes, "Better approximation of betweenness centrality.," in *ALENEX*, J. I. Munro and D. Wagner, Eds., SIAM, 2008, pp. 90–100, ISBN: 978-1-61197-288-7. [Online]. Available: http://dblp.uni-trier.de/db/conf/alenex/alenex2008.html#GeisbergerSS08.

[6] *LanguageTool*, https://languagetool.org/, Accessed: 2022-03-15.