Katchaguy Areekijseree and Sucheta Soundarajan
Syracuse University
Syracuse, NY USA.

Nov 22, 2018

Dear Editors of the Journal of Web Science,

We wish to submit an original research article, entitled "*Crawling Complex Networks: An Experimental Evaluation of Data Collection Algorithms and Network Structural Properties*" for consideration for publication in the WebSci18 Special Issue of the Journal of Web Science. This work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere. An earlier version of this work was published in the 2018 WebSci conference, and we were invited by Drs. Geert-Jan Houben and Victor de Boer to submit this extended version to the WebSci special issue.

In this paper, we consider the problem of collecting network data by crawling, and perform an extensive experimental analysis of network crawling algorithms. In contrast to existing work, we do not seek to propose a new algorithm or determine which existing algorithm is 'best', but rather, investigate how the structural properties of a network affect the performance of eight popular network crawling algorithms.

We have modified and extended our earlier work in the following ways:

1. In response to the WebSci reviews, we have considered both directed and undirected networks (as opposed to only undirected networks, as in our earlier work).
2. Our earlier work considered only one query response model, the 'complete' query response model, in which all neighbors of a node are returned in response to a crawler's query on that node. In response to the WebSci reviews, we defined and evaluated four additional query response models: 'partial', 'paginated', 'in-out', and 'out' responses. These different query responses are motivated by real-world applications. Full details of these query responses can be found in Section III.B.
3. We have updated all figures in Sections 5 and 6 as suggested by a reviewer. In particular, we show the performance of all the methods rather than only the best method of each group in Section 5. Moreover, we add ticks, labels and use a 'facet grid' as a reviewer suggested. In Section 6, we plot each set of results as a line graph, instead of showing them as a table, to make the results easier to understand.
4. We have extended our experimental analysis to account for the changes above. In general, our initial observations on undirected graphs under the complete query response model hold true for directed graphs and other query response models, but we describe some exceptions.
5. We have extended our 'Guidelines for Users' in Section VII to include both undirected and directed networks, and to incorporate the extensions above.

We have no conflicts of interest to disclose. Please address all correspondence concerning this manuscript to us at *kareekij@syr.edu* or *susounda@syr.edu*.
We are grateful for your consideration of this manuscript.

Sincerely,

Katchaguy Areekijseree and Sucheta Soundarajan