Paper Summary

Krishna Y.Kamath (UIN: 119001316)
kykamath@cs.tamu.edu
Date: 24 March, 2011

I. Paper Citation Information

- Title: The very small world of the well-connected [1]
- Authors: Xiaolin Shi, Matthew Bonner, Lada A. Adamic and Gilbert Anna C.
- Publication details: HT '08: Proceedings of the nineteenth ACM conference on Hypertext and hypermedia

II. Summary

A. Problem Statement

Most communication online happens through the social connections people have online. These communications result in the formation of social graphs which are huge in size. To understand the flow of information, design efficient algorithms to mine these graphs, we have to study the statistical and graph properties of the entire network. The properties that we need to understand might be degree distribution, clustering properties and the features of the network governing its growth and evolution. But the large size of social networks makes the calculation of these properties difficult.
One of the solutions to this problem has been identifying important or critical nodes in these networks and evaluating properties of these reduced graphs. But, there is no principled way study of how to reduce these graphs and ensuring the reduced graphs represent the quality of the actual graph. Having a accurate way to reduce a huge graph is important because it there are several online social networks that can be reduced without loss of quality. The reduced graph can be stored efficiently. The paper tries to build a simple principled method to build the reduced graph. Reduced graph is also called as synopsis of the actual graph.

B. Proposed Solution

The paper build the synopsis of the graph using important vertices, using which most of the properties of the actual graph can be identified without losing the actual graphs quality.
They paper first defines steps to identify the important nodes in the network. It defines 4 parameters based on which importance of nodes can be defined. Degree of a node is number of other nodes it is connected to. Betweeness of a node is the number of pair of nodes that have a path through the node. Closeness of a node measures the distance from other graphs. Pagerank is a method of measuring the importance of a node in comparison to its neighbors. The paper uses 4 datasets- erdos-reyni graph, buddyzoo graph, TREC and world wide web graph, to compare their analytical results empirically.
The paper first analyzes the actual network for the 4 properties mentioned before. They then build a subgraph of the important nodes and calculates these properties for this graph. The paper them compares the properties of the actual graph with the synopsis graph, built using important nodes. The paper then compares the quality of the compressed graph. It then analyzes 2 heuristic algorithms KeepAll and KeepOne. Theoretical analysis of

C. Results

After studying the graph they observe that a large graph can be reduced to a sub-graph of important nodes. They observe that different importance measures generate different sub graphs of varying density and topology. They identify different nodes as important, but inspite of these difference they agree in certain parts that are important for developing synopsis is a graph. Hence, they conclude that sub-graphs can be used to determine the properties of the large online social networks.

III. Critique

A. Strenghts

The area of online social networks is very interesting and the paper deals with the problem of understanding these networks. The huge size of social networks makes it difficult to analyze these graphs. The paper tries to build a sub-graph of the larger graph that reflects the properties of the larger graph. The problem they are trying to solve is important, hence the paper's problem and solution is relevant to its field of research.
The paper is very well written. The authors clearly define the problem they are are trying to address. They define, in advance, the various importance metrics use in the paper, like degree distribution, page rank, betweeness and closeness. They then perform theoretical analysis of their solution. The experiments they perform clearly show that their technique meets the goals desired by the paper.

B. Weakness

One of the places that, I thought, the paper could improve was in their experimental setup. The data-set that the paper uses to test its results are not very diverse. The size of the data-set is small compared to many actual social networks. Hence, testing their results on these larger graphs might have helped us understanding their techniques better.

IV. Future work

The results produced by the paper are interesting. Their work can be taken ahead in different ways. We can take their solution and apply that on larger and more real graphs and see if we observe results similar to this paper. Understanding if these techniques also work on larger graphs is important to the real world use of their techniques.

References

[1] X. Shi, M. Bonner, L. A. Adamic, and A. C. Gilbert, "The very small world of the well-connected," in *Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*, ser. HT '08. New York, NY, USA: ACM, 2008, pp. 61–70. [Online]. Available: http://doi.acm.org/10.1145/1379092.1379108