Referee report
on the PhD Thesis by Jithin K. Sreedharan
'Sampling and Inference in Complex Networks'
Referee: Nelly Litvak

The thesis has a broad but clearly defined research goal: inferring networks using only a small amount of information, which can be obtained on-line while crawling the network, and without knowing the network in advance.

The problem is of high practical importance. I was impressed by the balance achieved in the thesis between its broad scientific scope (from spectral analysis to degree-degree correlations) and its very focused unified goal. In my experience, this is a very exceptional combination.

Another strong point of the thesis is the ability of the candidate to invoke state-of-the-art knowledge from a broad literature. Chapter 2 builds on convergence and mixing time of Markov chains. Chapter 3 uses spectral smoothing and quantum random walks, and the recent theory connecting the spectrum of the graph with degrees of the vertices. Chapter 4 uses random graph models, as well as Bayesian statistics. Chapter 5 uses reinforcement learning, while Chapter 6 is based on the methods from the Extreme Value Theory.

The thesis consists of introductory Chapters 1 and 2 and content Chapters 3-6. Given the broad variety of methodologies used in the thesis, writing an introduction is a real challenge. I believe that the candidate did very well on introducing and motivating his research goals. The introduction to the methods is somewhat scattered, to my opinion. For example, the candidate presents the main concept from the extreme value theory in Chapter 1, devotes a complete Chapter 2 to random walks (which are indeed the main methodological tool in the thesis), while other methods used in the thesis are introduced in the corresponding content chapters. I believe that introduction to the methodology could be presented in a more coherent way. However, I understand the choices made, and this does not affect the quality of the content.

In Chapter 3, I liked the idea to evaluate the complete spectrum of the adjacency matrix A by estimating the extreme points of the function $\mathbf{f}_\theta$, which in fact represents a Fourier transformation of the exponent of A. The literature on finding eigenvalues and eigenvectors of A is very rich. In addition to the many references in the thesis, I suggest to look at the Arnoldi methods by Frahm and Shepelyansky:

Frahm, Klaus M., and Dima L. Shepelyansky. "Ulam method for the Chirikov standard map." The European Physical Journal B 76.1 (2010): 57-68.

Next interesting idea in Chapter 3 is the connection between complex diffusion and quantum random walk. Numerical results, based on the proposed methods, are very convincing.

The novel idea in Chapter 4 is introducing a so-called super-node, which can be created in a static or dynamic way. Then the average values of a function of a network can be estimated by running random walks on the network, where the super-node is viewed as a single node. The numerical results show that the average functions on networks can be estimated in this way with good precision

while crawling only a relatively small fraction of the network. Mathematically, the challenge is to prove the consistency of the proposed Markov-chain based estimators, which is accomplished in the thesis in both frequentist and bayesian framework. The methods are successfully applied to real-life networks and to the important null-models (configuration model and generalized random graphs). Interestingly, the real-life networks and the random graphs show very different behavior, which is another important finding in the thesis.

Chapter 5 looked to me as work in progress on Reinforcement Learning technique, which is in the spirit of the value iteration method in stochastic dynamic programming. This is an original technique, and the numerical results look promising.

Finally, Chapter 6 characterizes degree-degree correlations using a single parameter - the Extremal Index (IE), which the candidate proposes to evaluate, using empirical copulas. This approach of course requires a certain parametrization of the degree-degree correlations through the EI, which is only possible for a certain class of joint distributions of the degrees on the two ends of a random edge. Indeed, the candidate often uses a special bi-variate Pareto model. However, the idea is very interesting and possibly can be extended to a broader class of joint distributions.

I found the thesis very interesting and the developed methods original and promising. I recommend to award PhD dergree to Jithin K. Sreedharan.

Dr. Nelly Litvak,
Associate Professor, University of Twente
P.O.Box 217, 7500AE, Enschede, The Netherlands
tel: +31(53)4893388