

Title	時系列データに基づいた Scale Free Graph モデルに関する研究
Author(s)	森本, 真一
Citation	
Issue Date	2009-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/8101
Rights	
Description	Supervisor: 上原隆平, 情報科学研究科, 修士

Scale free graphs based on time sequential data

Shinichi Morimoto (0710069)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 5, 2009

Keywords: Blog data, scale free graphs, interval graphs, max-tolerance graphs.

Recently, scale free graphs and small world attract attentions since they can model many social networks including WWW and the internet. Because of their non-uniform structure, those graphs are considered to be able to model more various social networks than traditional Erdős-Rényi's random graphs having uniform structure.

Scale free graphs obey a law called “power law” and some models that are based on scale free graphs and obey this law are known. But analyses of these model's distribution of degrees are used to be done with differential equations. Therefore, analyses of graphs that are generated from the models are complex, and it is not easy to see the combinatorial structures of the graphs. It is important to simplify the analysis of the models to design algorithms for scale free graphs. Scale free graphs based on time sequential data called “scale free interval graphs” are proposed as a model which can be analyzed by relatively simple probabilistic and/or combinatorial arguments.

Scale free interval graphs are interval graphs. Interval graphs are a kind of intersection graphs. A graph $G = (V, E)$ is an interval graph if and only if G has an interval representation \mathcal{I} such that each vertex v corresponds to an interval I_v and two vertices u and v are adjacent in G if and only if corresponding intervals I_u and I_v share a common interval.

A scale free interval graph is an interval graph whose interval representation expresses time sequential data of biased living probability, and is a scale free

graph. This model is theoretical and probabilistic model. Unfortunately, this model does not fit the real networks. In scale free interval graphs, a vertex set corresponding to intervals that share an identical time induces a clique, and this property is not realistic.

We propose a new model of scale free interval graphs that has properties of real networks by using a blog data. The blog data are from Excite Japan Company Limited.

In this paper, we consider the blog data as an undirected graph: an (unique) ID in the blog data corresponds to a vertex and a link or a trackback corresponds to an edge. This graph has a lot of components. Therefore, we analyze a maximum connected component in the graph. We treat the blog data as intervals. The start point of an interval corresponds to the first update day of the blog and the end point of an interval corresponds to the last update day of the blog. As a result of analysis of blog network, the degree and the interval distributions show that the network is a scale free interval graph obeying power law.

Max-tolerance graphs can be regarded as generalized interval graphs, where two intervals I_i and I_j are joined by an edge in the corresponding graph if they overlap for an amount of at least $\max\{t_i, t_j\}$ where t_i is an individual tolerance parameter associated to each interval I_i . In this paper, we use max-tolerance graph in substitution for interval graphs. Using max-tolerance graphs, we can avoid the case that vertices sharing an identical time induce a clique. Thus we propose scale free graph models that have properties of real network.

In this paper, we propose three new network models, and make experiment to test the validity of these models. The first model is the deterministic model. The deterministic model decides whether two vertices u and v are adjacent under a certain condition. The second model is the probabilistic model. In this model, we analyze what is the main influence on the probability that two vertices are joined by an edge. The candidates are the length of intervals, the number of update times, the frequency of the updates, etc. Next we rearrange each edge with probability determined by one of the candidates, and check whether the resultant graph reconstructs the original degree distribution. The last model is the model which combines the probabilistic model and the preferential attachment model. The preferential attachment model is the most acceptable model among the known scale free graph models.

The results of the experiments of the three models are as follows. The de-

terministic model does not have scale free property. The probabilistic model has scale free property for large degrees, but does not obey power law for small degrees. The last model has scale free property, although the slope of degree distribution of the model is different from that in blog network.