Heuristic algorithm for determination of local properties of scale-free networks

Mitrovic, M (IPB)

01 February 2007

EGEE is a project funded by the European Commission
Contract number INFSO-RI-508833

The electronic version of this EGEE Publications is available
on the CERN Document Server at the following URL:
<http://cdsweb.cern.ch/search.py?p=EGEE-PUB-2006-039>

Heuristic algorithm for determination of local properties of scale-free networks

*Marija Mitrović **
Institute of Physics, Pregrevica 118, 11080 Belgrade
** mitrovic@phy.bg.ac.yu*

Introduction

- Complex networks are everywhere. [1,2]
- Many phenomena in nature can be modeled as networks:
  - brain structures
  - protein-protein interaction networks
  - social interactions
  - the Internet and WWW.
- They can be represented in terms of nodes and edges connecting them.
  Example: science collaboration network
  - node = individual scientist
  - edge = co-authorship of scientific article
- Important characteristics:
  these networks are not random; they have a structured architecture.
- Structure of different networks are similar:
  - all have power law degree distribution (scale-free property)
  - despite large size there is usually relatively short path between any two nodes (small world property).
- Global characteristics:
  degree distribution, clustering coefficient and the diameter.
- Local structure:
  frequency of subgraphs of given type (subgraph of order *k* is a part of the network consisting of *k* nodes and edges between them).
- There are different types of subgraphs of the same order (Fig.1).

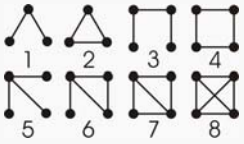

Figure 1: Type of subgraph up to order four.

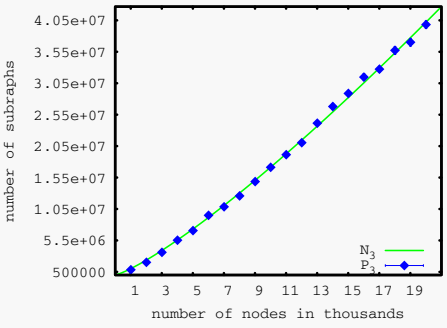

Figure 2: The number of subgraphs of type 3 estimated by the preferential approach for an incomplete 20% search.

The Model

- Barabasi-Albert algorithm: [1]
  - start with random network consisting of a small number (n_0) of nodes, mutually connected with probability *p*.
  - add a new node with *m* ($\leq n_0$) edges hat link the new node to *m* different nodes already present.
    The probability that a new node will be connected to node *i* is proportional to degree (number of links) of node *i.*
    After *t* steps this algorithm results in a network with $n=t+n_o$ nodes.
- We count the subgraphs of order up to four using two approaches, random and preferential.
- We chose a subset *S* containing a fraction of all nodes, and count all subgraphs of order three and four that contain at least one node from the subset.
- Random method: every nodes is chosen with the same probability, 1/*n.*
- Preferential method: the node is chosen with probability proportional to the degree of the node.
- Goal:
  - find ratios $a_i=P_i/N_i$ for the given network (N_i is a number of subgraphs of type *i* obtained by the exact counting, and P_i number of subgraphs found by heuristic search algorithm)
  - estimate the frequency of subgraphs in complex networks.

Motivation

- Global properties have been extensively studied.
- We investigate if two networks are similar by examining their local characteristics.
- The exhaustive search of subgraphs in a large network is computationally prohibitively expensive.
- We need an efficient heuristic approaches allowing counting of subgraphs by an incomplete search.

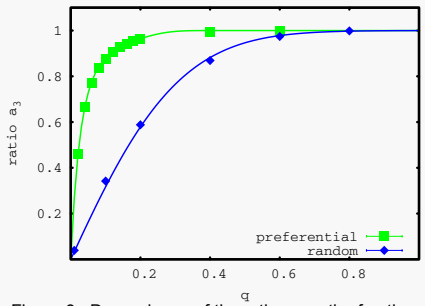

Figure 3: Dependence of the ratio a_3 on the fraction of searched nodes q, for preferential and random approaches.

Results and conclusion

- Here we report on the comparison of efficiencies of two heuristic algorithms [3], based on random and preferential choices of the subset of studied nodes.
- Both were applied to scale-free networks, generated using the Barabasi-Albert algorithm, to count subgraphs with up to four nodes.
- We find [3] that the preferential heuristic algorithm gives accurate estimates of local network properties. The random method, though simpler, is found to be less accurate.

References

1. Barabasi, A.-L., R. Albert, *Reviews of Modern Physics*,(2002) vol. 74, Issue 1, 47-97
2. N. Przulj,D. G. Corneil, and I. Jurisica, (2006) *Bioinformatics*, vol. 22, number 8, 974-980
3. M. Mitrovic, A. Belic, in preparation

Acknowledgments

This work was supported in part by the Ministry of Science and Environmental Protection of the Republic of Serbia, under project no OI141035. Marija Mitrovic has received support under FP6 project CX-CMS. The presented numerical results were obtained on the AEGIS e-Infrastructure whose operation is supported in part by FP6 projects, EGEE-II and SEE-GRID-2.