TARL (Topics, Aging and Recursive Linking)

Description

TARL (Topics, Aging and Recursive Linking) is a general process model developed by K. Börner, J. Maru and R. L. Goldstone which models the simultaneous evolution of author and paper networks. The model attempts to capture the roles of authors and papers in the production, storage and dissemination of knowledge. Information diffusion is assumed to occur directly via co-authorship and indirectly via the consumption of other author's papers. The model generates a bipartite evolving network which also incorporates aging in the paper citation network.

The model uses the simplifying assumption that there is a single level of specific topics. A fixed number of authors are generated each year. The set of authors are interlinked via undirected coauthorship relations. Papers are generated by these authors when they find a sufficient number of coauthors in their topic. The number of coauthors for a particular paper is also kept fixed for simplification. Authors "consume" papers before they "produce" a new paper. So authors and papers are linked via directed "consumed" links denoting the information flow from paper to author and directed "produced" links when a new paper is generated by these authors. The in-degree of a paper node refers to the number of references and the out-degree to its number of recieved citations. Some other simplifying assumptions that the model has are the following. Each author generates a fixed number of papers each year. Each paper has a fixed number of references. Each author and each paper have a fixed topic.

The model starts with an initial number of authors assigned to a fixed number of topics randomly. At every timestep a fixed number of authors are added. After certain years a fixed number of authors are deleted and are unable to coauthor with new authors due the finite lifespan of authors. Papers though once generated can be cited at any time. However to reflect the fact that more recent papers are being cited, the probability of citing a paper is modeled by a Weibull distribution. The Weibull distribution is over time and as age increases the probability of citing a paper goes down. So very old papers tend to get very few or no citations since they are in the tail of the distribution.

Pros & Cons

The model was successfully validated by a 20 year (1982-2001) data set of articles published in the PNAS (Proceedings of the National Academy of Sciences).

Applications

The model can be used to generate bipartite networks of coevolving authors and papers. It can be applied to other datasets with different aging distribution.

Implementation Details

The model has been implemented in Java. There are two input files which are provided. On clicking the TARL model, you have to choose these files. The files can be found in the directory /sampledata/Network/TARL. They are iniscript.tarl and agingfunction.txt.

The iniscript.tarl contains all the input parameters. The model can be started with or without topics. The first line of iniscript.tarl indicates that. Aging can either be enabled by setting the counter to 1 or disabled by setting the counter to zero. Then the following parameters have to be supplied. StartYear, EndYear, number of authors in StartYear, number of papers in StartYear, the maximum age of authors, the number of topics, number of authors to be deleted per 10 year(s), growth in number of authors per 1Year(s), number of publications per year, number of co-authors, number of papers read each year, number of papers cited each year, number of papers produced each year and the number of references considered.

The second file is the agingfunction.txt. This file gives a probability distribution that the program reads. The file consists of a single column (the dependent or y values). The independent (or x) axis is considered to be the age which starts from 0 and goes to 21. The distribution is a Weibull distribution with shape parameter 2 and scale parameter 3.

Usage Hints

You may generate small networks starting with a small number of authors and papers and visualize a bipartite network or two unipartite networks. The two input files have very specific formats so do not change the formats of the files provided. You can however change the input values in the iniscript.tarl file and generate networks with various numbers of authors and papers. The agingfunction.txt can be changed by a distribution with a different scale factor but the distribution should be properly normalized and should be in the same format, i.e. one column with the y values only.

Acknowledgements

The JAVA implementation is by Jeegar T. Maru and the algorithm has been integrated by Ramya Sabbineni. Documentation was compiled by Soma Sanyal.

References

Börner, K., Maru, J. T. and Goldstone, R. L. (2004). The simultaneous evolution of author and paper networks. PNAS. 101(Suppl_1):5266-5273.

Links

- Source Code

See Also

⊘ The license could not be verified: License Certificate has expired! Generate a Free license now.