

# Data Management Plan

To date, research on algorithms for managing temporal networks has included very little in the way of comparative empirical evaluations of competing algorithms. A major emphasis of the *Temporal Reasoning Lab* for this project will be to carry out comprehensive, rigorous, and reproducible empirical evaluations and comparisons of both pre-existing algorithms for managing temporal networks and new algorithms developed as part of the project. To enable the empirical evaluations requires not only implementing the candidate algorithms, but also creating software for randomly generating benchmark problems that model the features from a wide range of real-world applications. To ensure transparency and reproducibility, all algorithms, implementations, benchmark software, experimental parameters, collected data, and statistical analyses will be thoroughly documented and placed within a repository that will be accessible to the wider research community, thereby enabling all empirical results to be independently validated.

## 1. Types of data.

- A. Computer programs (plain text files), written in standard computer programming languages such as Lisp (.lisp), Java (.java) or Python (.py), that are implementations of temporal reasoning algorithms and software for generating suites of benchmark problems.
- B. Data generated by empirical evaluation of temporal reasoning algorithms: log files (.txt) and statistical analyses (.csv, .txt).
- C. Documentation for software: plain-text files (.txt) or PDF files (.pdf).

## 2. Standards for format and content.

### A. Naming conventions.

- i. Implemented Algorithms: *ALG:Net\_Alg\_Auth\_Ver.Ext*, where *Net* specifies the type of temporal network (e.g., STN, STNU, etc.), *Alg* specifies the type of algorithm implemented (e.g., DCCheckingMorris2006), *Auth* specifies the author of the program, *Ver* is the version number, and *Ext* is the file extension for the programming language (e.g., lisp, java, py).
- ii. Log Files: *LOG:Net\_Auth\_Ver\_Date.txt*, where *Net*, *Auth* and *Ver* are as described above, and *Date* specifies the date the data was collected in mm-dd-yyyy format.
- iii. Statistical Analyses: Same as for Log Files, except prefixed by *STAT* instead of *LOG*.
- iv. Software for Generating Benchmark Problems: *BENCH:Net\_Auth\_Ver.Ext*, where *Net*, *Auth*, *Ver* and *Ext* are as described above.
- v. Documentation for software: *DOC:Label.Ext*, where *Label* is the filename of the Software being documented, and *Ext* is one of .txt, .tex or .pdf.

### B. Format. Log files and statistical analyses are text files (.txt or .csv).

### C. Content. The log files include a preamble that includes the date (mm-dd-yyyy) and time (24-hour clock) of the experiment, the types and versions of the software used, makes and models of all computers; the environmental configuration; and the seeds for any random number generation. Once the statistical analyses are completed, the raw data need not be saved, since it can be re-generated from the information in the preamble.

## 3. Sharing and access policies.

### A. Access sites.

As data are made available, they will be web-discoverable at `cs.vassar.edu`. The site will be

tagged by this NSF proposal title, NSF proposal ID, the name of the PI, and keywords. The server site, `cs.vassar.edu`, is maintained by the Systems Administrator for the Computer Science Department at Vassar College, who maintains a firewall and performs daily backups, kept off-campus at a secure site.

- B. Publication, during and after the grant. All data and files of the types covered by this Data Management Plan during the grant will be made available first to reviewers of submitted manuscripts. Once a paper has been made available publicly, all such data and files will be made available upon request, via the journal's or society's web page and `cs.vassar.edu`.
  - C. Open access for publications. All peer-reviewed or juried conference papers are made available via self-archiving ("green OA") per NSF requirements (document NSF15052). As required, the final papers will be deposited at the Department of Energy's PAGES system, and at `cs.vassar.edu`. Journals or conferences that forbid public access within one year of publication will not be used.
4. Re-use, re-distribution, and derivatives policies. All data, files, and papers are made available without restriction at the repository at `cs.vassar.edu`. The only exceptions are: (1) in the case of OA publications with copyright restrictions, such restrictions will be posted on the item description in the repository; and (2) computer programs and documentation files will contain a preamble that must be included in any products derived from them.
  5. Storage, preservation and archiving after grant ends. All data, files, and papers of the types covered by this Data Management Plan will be stored, preserved and archived on the servers at `cs.vassar.edu` maintained by the Computer Science Department at Vassar College.
  6. Roles and responsibilities. The PI bears primary responsibility for data as it is generated, analyzed, and submitted for publication. Since publication triggers the public access of the data, the PI bears immediate responsibility for making data available as described above within three months of the date of initial publication, unless the policy of the journal or conference mandates a 12-month embargo as per OA green. Should the PI take a full-time position at another institution, he still bears this responsibility. Once data and publications are deposited at `cs.vassar.edu`, the CS Department's Systems Administrator bears the primary responsibility for the storage, access, archiving, and maintenance of the data and publications.