Data Management Plan

In this project, we will collect information and plan the steps for creating a community infrastructure for analyzing multilayer networks. All data will be available and accessible through a publicly-accessible project webpage.

1. Types of data to be produced to be collected and generated in the course of the project:

   - Software and algorithms included in the prototype of the community infrastructure.
   - Documentation for the community infrastructure.
   - Recordings of online lectures on multilayer networks, as well as presentations from the workshop
   - Report on the requirements of the community for a multilayer network analysis tool

2. Standards to be used for data and metadata format and content

   We plan to use existing data and metadata formats for any data we consume (such as graphs and matrices in text file formats). If we change the representation format while processing the data, we will provide the necessary programs to convert from the original data source.

3. Physical and/or cyber resources and facilities (including those supplied by third parties) that will be used to store and preserve the data after the grant ends

   We plan to use compute facilities available at our host institutions for developing and prototyping algorithms. All software will be developed under version control (Git, Subversion) and made available on publicly-accessible platforms such as SourceForge or GitHub.

4. Policies for access and sharing

   The community infrastructure prototype will be made available under the open-source BSD license. Educational material will be made available for download under the Share-alike Creative Commons License. The license gives people the right to share and use both the work and any derivatives, as long as they are distributed under the same license. Digital Object Identifiers to publications will also be made available through the project website. Wherever possible, we will share cleaned and processed data corresponding to snapshots of the full dataset used in research evaluations. If the data source does not permit direct sharing of processed data, we will provide software and scripts that automate the process.

   We will not host any networks or data that are sensitive to privacy issues. We anticipate that most of the data obtained from our collaborators will be available for distribution, without any constraints. The projects generated as part of the workshop or student participation in the research and related courses will be made available after obtaining permission from the student.

5. Policies and provisions for re-use, re-distribution, and the production of derivatives

   We will adhere to the licensing policies of any third-party tools and data we use. However, the standalone software developed by us can be freely re-used and re-distributed.

6. Plans for archiving data and other research products, and for preservation of access to them after the grant ends

   For short-term storage (i.e., during the project duration) of research data, we plan to use a dedicated workstation at Penn State, with cloud backup using Box or Dropbox. We plan to use the Penn State Libraries and IT Services' ScholarSphere research repository service for longer-term storage (i.e., after the grant ends) of research data generated through this project. In addition to preservation and secure backup, the service offers metadata search and citation support.

2120393

7. Roles and responsibilities of all parties with respect to the management of data after the grant ends

The PIs will ensure access to all types of data generated through this project for at least 3 years after the grant ends. Data will be publicly available via Penn State's ScholarSphere platform as well as publicly-accessible platforms such as GitHub, SourceForge, and Box. We hope that making research data available on public platforms that are not associated with our primary institutions (such as GitHub, SourceForge, and Box) will guard against any unanticipated departure of key personnel from the project.

2120393