



DATA MANAGEMENT PLAN

Project Number: 840922

Project Acronym: TEAMS

Project Title: Modelling Trust-based Evolutionary Dynamics in Signed Social Networks

Data Management Plan

Name of Fellow: Fei Hao

Department: Department of Computer Science

Name of Supervisor: Geyong Min

1. Data Summary

1.1 Background

The TEAMS project participates in the Horizon 2020 Pilot on Open Research Data. TEAMS's Data Management Plan (DMP) describes the types of data that will be utilized and generated during the project, the standards that will be used, how the data will be exploited and shared for verification and reuse, as well as the preservation and storage of data.

The provisions for open access to research data are described in Grant Agreement Arts 29.3 *Open access to research data* and 39.2 *Processing of personal data by the beneficiaries*. This document follows the template for the H2020 Data Management Plan v1.0 (13/10/2016).

1.2 Objectives of the TEAMS project

TEAMS aims to carry out the innovative research to develop the mechanisms for trusted communities detection and evolutionary dynamics in signed social networks (SSNs) in order to support personalized goods recommendation. Two specific objectives are as follows.

Objective 1: Establish the new model and develop innovative algorithms for detection of the trusted communities. By investigating the Formal Concept Analysis (FCA)-based representation model for SSNs, this project plans to establish the equivalence relation between the trusted communities and the formal concepts which are the basis of the FCA-based trusted communities' detection. Based on this model, we intend to implement (1) an FCA-based trusted community detection algorithm; (2) an FCA-based diversified trusted community detection algorithm.

Objective 2: Establish the evolutionary model and develop the innovative algorithms for the trusted communities. Based on the above extracted trusted communities, we will analyse the impacts on social balance theory for the evolutionary dynamics of the trusted communities. Further, the trusted community evolutionary model and algorithm will be developed.

During this fellowship, we aim to formulate the following novel theory and further develop a prototype system based on the theory.

Theory: Aiming to efficiently detect the trusted communities and diversified top-k trusted communities from SSNs, the equivalence theory which maps the trusted communities in SSNs domain to the formal concepts in FCA domain will be presented; this equivalence theory will provide a critical solution for bridging the gap between social computing and soft computing disciplines and further address the social issues, such as rumor control and fraud information inhibition with FCA methodology.

System: A personalized recommendation system for goods will be designed by developing an extensible lightweight software component or plug-in for the system. This unique recommendation system can be used for e-commerce, social marketing and social advertising and so forth.

1.3 Types of data utilized and generated

TEAMS will conduct social computing research by using the existing relevant datasets which are available in the public, such as Epinions, Wikipedia and Slashdot as shown in Table 1. Note that these datasets are released and maintained by SNAP group at Stanford University. SNAP opens the usage permissions to all researchers. However, the raw files as such will not be archived for reasons of ownership and copyright. For the data that is already available in the public domain, references to the original data will be made in the publication.

Table 1. Types of data utilized and generated in TEAMS

Data	Description	Source	File format	Size
Public Dataset				
Epinions	This is a who-trust-whom online social network of a general consumer review site Epinions.com . Members of the site can decide whether to "trust" each other. All the trust relationships interact and form the Web of Trust which is then combined with review ratings to determine which reviews are shown to the user.	SNAP network analysis platform (http://snap.stanford.edu/)	Text document Txt (*.txt)	1.55 MB
Wikipedia	The network contains all the Wikipedia voting data from the inception of Wikipedia till January 2008. Users can make "positive vote" or "negative vote" for the promotion of adminship.	SNAP network analysis platform (http://snap.stanford.edu/)	Text document Txt (*.txt)	283 KB
Slashdot	It is a technology-related news website known for its specific user community. The website features user-submitted and editor-evaluated current primarily technology oriented news. In 2002 Slashdot introduced the Slashdot Zoo feature which allows users to tag each other as friends or foes. The network contains friend/foe links between the users of Slashdot. The network was obtained in February 2009.	SNAP network analysis platform (http://snap.stanford.edu/)	Text document Txt (*.txt)	3.1MB
Generated Dataset				
Trusted Communities	Trusted community is a special community where all the links are with the trust relations.		Text document Txt (*.txt), (*csv)	WP2
Diversified top-k Trusted Communities	Top-k trusted communities that have the diversification features, <i>i.e.</i> , the overlapping among these communities are minimized.		Text document Txt (*.txt) (*csv)	WP2
Frequent Trusted Communities	The trusted communities which have the higher appearance frequency over the time.		Text document Txt (*.txt) (*csv)	WP3
Constant Trusted Communities	The trusted communities which are not changed over the time.		Text document Txt (*.txt) (*csv)	WP3

1.4 Data utility

The data utilized and generated by TEAMS might be useful for the broader scientific community for the sake of validating results presented in TEAMS's scientific publications, as well as for future social computing research, especially for link prediction and community detection in signed social networks.

2. FAIR data

TEAMS complies with the guiding principles for scientific data management outlined by the H2020 programme, which ensure that the research data is **Findable, Accessible, Interoperable** and **Reusable** (FAIR).

2.1. Making data findable (F)

Persistent Digital Object Identifier (DOI)

Upon archiving in a data repository, each dataset will be assigned a unique DOI or URL. As a general rule, TEAMS's datasets will be archived in the fellow's hosting institutional repository, Open Research Exeter (ORE)¹/ as well as TEAMS project website². This can be included as part of data citation in publications, allowing the datasets underpinning a publication to be identified and accessed. DOIs will also be linked with appropriate publication records in ORE, the University's institutional repository, to enhance visibility of datasets.

Metadata

The metadata for each dataset will include information on the following: creator, access conditions, generation methods, time references, structure and organisation of data files, file formats, variable names, labels and descriptions of variables and values. Importantly, the search keywords to the list of metadata will be added for better enhancing the findability of the data.

Naming conventions

TEAMS datasets will be re-named with project title (TEAMS), creator/owner of the dataset will be provided. The naming rule is as follows:

TEAMS-00-datatype [i.e., TEAMS-01-Epinions]

Here, "01" represents the ordering number of dataset.

In addition, the generated data files will be named as: Trusted Communities: **TEAMS-TComm**, Diversified top k Trusted Communities: **TEAMS-DKTCComm**, Frequent Trusted Communities: **TEAMS-FTComm**, or Constant Trusted Communities: **TEAMS-CTComm**.

2.2. Making data openly accessible (A)

Data repository

After the end of the project (31/1/2022), and as soon as project results have been published in the form of peer-reviewed publications, all generated data from TEAMS, including metadata documentation and code, will be deposited in a ORE where research data can be securely preserved for the long-term. Data deposited into ORE are highly discoverable and will be actively curated, which may involve changing the format of the

¹ <https://ore.exeter.ac.uk/repository/>

² <https://fhaocs.github.io/TEAMS/>

data to ensure long-term accessibility and re-usability. Metadata about datasets will be publicly searchable and discoverable and will indicate how and on what terms the dataset can be accessed.

Website

The TEAMS project website (<https://fhaocs.github.io/TEAMS/>) will make reference to the data (including metadata) generated which is available for further exploitation and link to the data repository where it is stored. It will also increase the visibility of the datasets.

2.3. Making data interoperable (I)

Standards and formats

The data to be deposited and made available for other researchers will be archived in standard formats and in compliance with available open software applications to allow for data exchange and reuse. The file formats for different kinds of data are listed in Table 1. For better interoperability of the data, a readme file explaining the data will be created.

2.4. Increase data re-use (R)

Licencing

Data will be made available under a CC licence, which manages copyright and the terms and conditions for access and reuse. A CC licence requires attribution of the licensor (data supplier) when the use involves public sharing.

After the end of the project (31/1/2022), and as soon as project results have been published in the form of peer-reviewed publications, TEAMS data will be deposited and made available for third parties. The moment, at which the data will be made available, any conditions for re-use by third parties and embargoes after the end of the project will be specified in a subsequent version of this DMP.

3. Allocation of resources

3.1 Costs

Costs related to data management are expected to be limited and will be covered by the individual beneficiary's institutional unit costs. The fellow will be responsible for setting up a budget for data gathering/collection with the help of local research and/or financial support staff.

The costs will be discussed with the beneficiary (supervisor and research manager/financial officer) and the Coordinator and Project Manager, who will make sure that the provisions of the TEAMS Grant Agreement and this DMP are followed. The potential value of processing such data with the purpose of making them available for re-use by other researchers must be estimated as higher than the costs in terms of the work load, which will depend of the amount/size of the data, and sensitivity/confidentiality issues.

Costs related to the administrative coordination of data management and deposition in line with this DMP will be covered by project funds at network level administered by the coordinator.

Costs for the long-term preservation of data will not occur, as the archiving in the ORE.

3.2 Project responsibilities

The host Prof. Geyong Min will be ultimately responsible for data management of this project. The fellow will be responsible for organization and storage of the data as it is produced during the project. The College of Engineering, Mathematics and Physical Sciences' IT staff will manage the school's server, where the data

will be stored. The University's library staff will be responsible for management of the institutional repository, ORE.

4. Data security

TEAMS's fellow is employed at University of Exeter which offers several options for secure storage of data. Throughout the project, in the data gathering and analysing phase, individual researchers will store data in their personal password-protected spaces on secure university servers, which provide appropriate back-up solutions.

At the end of the project, the data will be deposited for long-term storage in certified repositories, such as the Exeter ORE data archive and TEAMS project website which is host by Github.

5. Ethical aspects

There are no ethical issues concerning TEAMS project.

6. Other issues

No, TEAMS does not use other national/funder/sectorial/departmental procedures for data management.