Plan Overview

A Data Management Plan created using DMPonline

Title: Evolving models for dyadic network data to represent and test complex theories

Creator: Terrence Jorgensen

Principal Investigator: Terrence Jorgensen

Data Manager: Terrence Jorgensen

Affiliation: University of Amsterdam (Universiteit van Amsterdam)

Funder: Netherlands Organisation for Scientific Research (NWO)

Template: Data Management Plan NWO v1

ORCID iD: 0000-0001-5111-6773

Project abstract:

Psychological theories of how individuals think, feel, and behave within social contexts (e.g., families, classrooms, work environments) are often too complex for current statistical models to represent. Treating individuals as discrete, isolated entities ignores information when studying inherently dyadic phenomena (e.g., dating, peer influence, bullying, family/classroom dynamics). I will develop structural network modeling to combine the advantages of generalized structural equation modeling (SEM), the social relations model (SRM), and social network analysis (SNA) models. The SRM can simultaneously estimate dispositional, situational, and interactive influences on behaviors or perceptions, but only recently have limited ad-hoc attempts been made to extend SRM beyond merely describing to actually explaining interpersonal psychosocial processes in the context of real human interactions, such as how peer conversations affect body image, disordered eating, or bullying. My project will extend SRM and SNA to account for measurement error in multiple indicators at various levels of measurement (binary, ordinal, continuous) and simultaneously evaluate theories (e.g., mediation, growth trajectories) at the level of individuals, relationships, and groups. SEM allows complex psychological theories to be both easily communicated via intuitive path diagrams and rigorously tested against empirical data. Recent publications reveal researchers' explicit demand for a generalized method to apply SEM to dyadic network data, which my project will accomplish. For example, the international antibullying program, KiVa, has been evaluated with multilevel SEM to test student-level and classroom-level effects of the program on bullying and victimization. Yet some observed indicators were aggregations across multiple interpersonal observations, which could still contain important information about whether KiVa works at the dyad level (on specific victim-bully relationships). My project will allow researchers to disentangle context-specific relationship effects from individuals' general interpersonal perceptions/behavior, consequently leading to more effective targeted interventions (e.g., prevent bullying particular individuals), and benefit behavioral scientists worldwide with easy-to-use open-source software.

ID: 33866

Last modified: 20-11-2018

Grant number / URL: 016.Veni.195.457

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Evolving models for dyadic network data to represent and test complex theories - 1 General information

1 General information

Name applicant and project number

Terrence D. Jorgensen
Project number: 016.Veni.195.457

Did you ask for support of the datamanagement support office of your institution?

- Yes

Evolving models for dyadic network data to represent and test complex theories - 2 Description dataset

Description of the data

Describe the data that will be collected/generated within the project.

No human-subjects data will be collected within this project. The only type of data that will be generated are simulated data for Monte Carlo simulation studies. Monte Carlo studies are used to evaluate how statistical methods behave under various known conditions. This is relevant to my project because I am developing a new family of statistical models to handle complex data, and I must assess how well they work.

Specify the type and format of the data.

Pseudorandom data will be generated via the open-source software package R, which is available free of cost to the world community of researchers. Results can be reproduced by anyone on any computer using plain-text syntax files that I will make available on my Open Science Framework account, along with metadata (also in plain-text files) to explain how the files can be used for reanalysis or for learning purposes. Storage of the simulated data is therefore unnecessary.

Evolving models for dyadic network data to represent and test complex theories - 3 Data storage during the project

Volume data

What is the volume (MB, GB, TB) of the data and where will the data be stored?

My project data will be stored in my "H:" network drive, provided by my institution's ICT department for every employee. Plain-text files take very little storage space, so I expect the entire project to require only 1-2 MB of storage capacity for the syntax files used to simulate data for my Monte Carlo studies. The "H:" network drive has a capacity of 10 GB, so no additional costs will be necessary for data storage.

Storage capacity

Is there currently sufficient storage capacity during the project?

- Yes

Will the data be backed-up regularly during the project? Who is responsible for this?

- Yes

I automatically backup all of my work-related files using SURFdrive (https://www.surfdrive.nl/).

Costs

What are the expected costs? Please specify these and state an amount that is as realistic as possible.

As an employee of the University of Amsterdam, I have access to a laptop for work (with my "H:" network drive mapped onto it) and a free SURFdrive account for automatically backing up my data (up to 250 GB, well beyone the capacity required for my project). There are no additional monetary costs associated with storing the data required for my Monte Carlo simulation studies. The creation of metadata will take considerable time, but that is a regular part of such research, and much of it is redundant with the details that are included in a publication of the Monte Carlo study's results. Furthermore, developing software to implement the models I develop is a component of my project, so the metadata are also largely redundant with writing documentation for my software, so that applied researchers can learn to use it. Because the data are simulated, there is no security risk.

How will these costs be covered?

There are no additional costs associated with storing the data required for my Monte Carlo simulation studies.

Evolving models for dyadic network data to represent and test complex theories - 3 Data storage after the project

Repositories

Specify in which trusted repository the data will be stored after the project.

Upon completion of my project, the project files will be permanently deposited in the University of Amsterdam's UvA/AUAS figshare repository. Additionally, data for each component of my overall Veni project will be stored on my Open Science Framework account, associated with its own project, complete with description, metadata, and PsyArXiv preprints of the manuscript(s) borne out of that project.

If the data will not be stored in a trusted repository how will the data be made findable, accessible and reusable?

Persistent Identifier

Will a persistent identifier be used to make the data findable?

Yes, every project on the Open Science Framework (OSF) has a digital object identifier (DOI) associated with it. Any publications resulting from this Veni project will include the DOI and permanent link to the associated OSF project where they can find the (meta)data. Using UvA/AUAS figshare to store all data from my completed project, every published item is assigned a DOI as well.

Period of storage

For how long will your data be archived?

I will store my project data in both the OSF and UvA/AUAS figshare for a minimum of 10 years. Each component of my Veni project will have a permanent home on the OSF. Even if they do run out of funding, they have secured enough funds to guarantee the files will continue to be hosted for at least an additional 50 years (see their FAQ page).

Costs

What are the expected costs? Please specify these and state an amount that is as realistic as possible. How will the costs be covered?

As an UvA employee, I can use UvA/AUAS figshare without any additional costs. Furthermore, the OSF hosts files free of cost.

Evolving models for dyadic network data to represent and test complex theories - 4 Standards and metadata

Standards and metadata

How will the data be documented? What metadata standard will be used to make the data accessible and reusable?

UvA/AUAS figshare uses the metadata standard Dublin Core, and OSF recommends using the existing standard of the discipline, so I will use Dublin Core as well. For each Monte Carlo study completed in my Veni project, I will create a plain-text README file that describes how to use the plain-text R-syntax files to

- reproduce the simulated data
- analyze the simulated data
- collect the Monte Carlo results into a single file, which can be exported to plain-text formats (e.g., CSV files) that can be imported into any general software package (e.g., R, Stata, SAS, SPSS, Excel, etc.)
- (when applicable) reproduce the analysis of Monte Carlo results reported in the associated manuscript

For the software I develop, I will generate "vignettes" using a standard format for including detailed documentation in R packages: http://r-pkgs.had.co.nz/vignettes.html

Evolving models for dyadic network data to represent and test complex theories - 5 Making data available

Sharing of data

Are the data, or part of these, available for reuse after the Open Access project?
If not, please explain why the data are not suitable and/or available for reuse.

- Yes, immediately

The Open Science Framework not only hosts a database of files associated with each project, but also provides a Wiki for each project that provides a detailed description and allows each project (and its associated data) to be easily searchable. For example, using a Google search of "osf jorgensen bayesian fit indices" returns one of my recent projects as the first search result. Because I use only open-source software to conduct my research, the data will be available to (and usable by) any researcher on the planet who has an internet connection. I provide these materials under a Creative Commons license (CC-By Attribution 4.0 International), so the only restriction is that anyone using th data must cite/attribute the source of the data (i.e., my project). Because science is a cumulative endeavor, researchers who access my project files can benefit either by (1) learning how to apply my methods to their own *real* data, or (2) design follow-up Monte Carlo studies that extend my research by testing the models I develop under different conditions.

If data are only made available after a certain period then please state the reason for this. If part of the data cannot be made (directly) available then please state the part concerned

Question not answered.

Are there any conditions for the reuse of the data?
If so, are these conditions defined in a consortium agreement?

For projects hosted on the OSF, as well as the project data stored, shared, and published on UvA/AUAS figshare, I will use a Creative Commons license (CC-By Attribution 4.0 International). For the open-source software I design, I use the GNU public license (version 3), which follows the philosophy that users can freely use, share, and change the software to suit their needs. This is quite important when developing software for cutting-edge research, because unforeseen needs arise on a regular basis.