FastDAS – User Guide

Installation

Installing Java

For most users, Java will be pre-installed on your system – FastDAS has been tested on Java 1.7 and higher. If required, it is available from https://java.com/download

The instructions here assume that you have Java added to your system's PATH environment variable, and this is required to run FastDAS with logging enabled – see https://java.com/en/download/help/path.xml for details.

Installation

Installing FastDAS

1. Extract the zip file to a location on your computer.

Installing required libraries (glpk-4.55)

FastDAS requires access to glpk and its Java bindings (glpk_java). From the glpk_java install instructions at (http://glpk-java.sourceforge.net/gettingStarted.html):

Windows

1. Download the current version of GLPK for Windows from https://sourceforge.net/projects/winglpk/.
2. The filename for version 4.55 is winglpk-4.55.zip. Unzip the file. Copy the glpk-4.55 folder to "C:\Program Files\GLPK\".
3. Check installation of GLPK by running the command
4. Make sure that glpk_4_55.dll and glpk_4_55_java.dll are accessible under your system's PATH environment variable.

Linux (Ubuntu – 14 LTS or later)

1. Install OpenJDK 8*:
```
sudo apt-get install openjdk-8-jdk
```
2. Verify that the default java is OpenJDK 8.
3. Install glpk-utils:
```
sudo apt-get install glpk-utils
```
4. Install libglpk:
```
sudo apt-get install libglpk-java
```

Alternatively, install from source. Linux users should note that the provided FastDAS jarfile was compiled using JDK 1.8.0_u25 – if you are using the OpenJDK or a different Java runtime environment, you will need to recompile FastDAS.

VirtualBox VM

A VirtualBox VM running FastDAS under Ubuntu 15 is available on request from the authors.

OSX

OSX instructions are provided as a guideline only – OSX is not supported.

Using the method originally described by Timothy H Chung of the Naval Postgraduate School, Monterey, CA, USA at https://wiki.nps.edu/pages/viewpage.action?pageId=113606659 (an excellent guide for troubleshooting this process is available at this page).

1. Make sure that Xcode is installed from the App Store.
2. Install Macports (following instructions at http://www.macports.org/)
3. Install required packages:
    a. `sudo port install glpk`
    b. `sudo port install swig`
    c. `sudo port install swig-java`
4. Download and unzip glpk for java from the project site at http://sourceforge.net/projects/glpk-java/
5. Set the JAVA_HOME environment variable in ~/.profile by opening that file in a text editor (e.g. `nano`) and adding the following line:

```
export
JAVA_HOME=/System/Library/Frameworks/JavaVM.framew
ork/Versions/CurrentJDK/Home
```

6. Set the following environment flags in ~/.profile:

```
export CPPFLAGS="-
I/System/Library/Frameworks/JavaVM.framework/Head
ers -I/opt/local/include"
export SWIGFLAGS="-
I/System/Library/Frameworks/JavaVM.framework/Head
ers -I/opt/local/include $SWIGFLAGS"
```

7. Run configure in the glpk-java source directory (with `./configure`).
8. Run make and check for errors (with `make && make check`).
9. Run make install (`sudo make install`).

Running FastDAS

FastDAS can be loaded by either double clicking the JAR file, or with logging enabled by opening a command prompt and using the command:

```
java –Xmx4g –jar FastDAS_<version>.jar
```

The -Xmx4g argument allows a maximum of 4Gb of RAM – make sure that you have this available (this is necessary for running large (Reactome-size) analyses). You may need to specify a java.library.path variable: under OSX, this would look like:

```
java -Xmx4g -Djava.library.path=/usr/local/lib/jni –jar
FastCompare.jar
```

Instantiating an experiment

A note on experimental design

Models within FastDAS should have a set of feasible state assignments which correspond to valid biology, and a set of infeasible state assignments which do not – in other words, where the solver fails to find an answer, it should be because the phenotype being described is impossible given the reaction network provided – either this situation directly contradicts the recorded knowledgebase, or the situation is not biologically plausible.

If you map information into the network, be careful that it is logically consistent. FastDAS is not a time-parameterised modelling technique, so statements that make sense when modelling reaction rates in terms of concentrations do not within the FastDAS framework.

For example, mapping species A as inactive and B as active in the reaction A -> B makes sense in a time-parameterised model if A is consumed when B is activated or produced. It causes logical inconsistency in FastDAS however; A must have been active for B to have activity (assuming no other reactions result in activation of B). This formulation requirement results from a lack of kinetic information available to describe these reactions.

Generating required files

FastDAS operations require some combination of three files:

1) A BioPAX Level 3 file describing a reaction system of interest in the format used by Reactome.

   This file will either be curated or not curated. Aspects of Reactome's annotation are used by the AutoCurator module in the curation process – for example, Reactome annotates its meta-entities with 'Converted from EntitySet in Reactome', whereas BioPAX versions of Panther Pathways/KEGG/etc use other annotations.

2) A pathway exclusion file:

   Choosing 'Generate pathway list' from the 'Generate' menu in the main FastDAS window creates this file. This file consists of the displayNames of Pathway objects in the BioPAX Level 3 file with one name per line. These pathways will be removed from the model along with their sub-pathways. Any reactions they contain will be removed unless they feature in a pathway that is being retained in the model; any subsequently orphaned species will be removed as well.

3) A species list:

   Choosing ‘Generate species list’ from the ‘Generate’ menu in the main FastDAS window creates this file. This file contains one species per row, and its format and use is described in detail in the next section.

Specifying an experiment – the species file

The species file is a tab-separated text file meant to be edited in a spreadsheet – if using a text editor or programmatically editing the file, please substitute in an appropriate set of tab separators.

The file consists of eight required columns, which are not labelled in the species file itself.

Column A is the object’s class name. It can be used to sort species by type – ‘class org.biopax.paxtools.impl.level3.ProteinImpl’ corresponds to proteins in the model. It should not be edited.

Column B is the RDF ID of the species, as recorded in the BioPAX Level 3 file, and can be used for cross-referencing to that file. It should not be edited.

Column C is the display name of the species, as recorded in the BioPAX Level 3 file. If there is no display name for this species, its RDF ID is used instead. It should not be edited.

Column D is the subcellular location of the species, and should not be edited.

Column E is the participation number of the species (*i.e.* how many reactions it participates in). It is not editable.

Column F is the network location of the node – INPUT, OUTPUT or INTERNAL. This is also not a directly user-editable property.

Column G is a flag that determines whether a species is retained in the model or not – it either has a value of RETAIN or REMOVE, and is meant to be edited by the user.

Column H is the reference model specification. To work with a generic Reactome model with no species activities specified, leave this column blank. Otherwise, set species activities to be one of ACTIVE, INACTIVE or FREE.

Column I onwards specify states for pairwise comparison against the state specified in Column H. Again, set species activities to be one of ACTIVE, INACTIVE or FREE.

Autocuration

1) From the file menu, select ‘Autocurator’. Model, pathway and species file must be loaded in the resulting dialog box.
2) Click ‘OK’. Autocuration can take several hours and a considerable amount of computational resources depending on the network size and complexity.

Running simulations

1) From the file menu, select 'Run Qualitative Logic System'. The below dialog box will appear.

2) Specify curated model, pathways and species file locations using their 'Browse' buttons.

3) Select an optimisation mode – either *minimize* or *maximize*. The majority of models will use *maximize – minimize* instances will frequently show only the reactions which individual species directly participate in, as they will toggle off as many inputs to the system as possible.

4) Select an objective function – either *inputs*, *outputs* or *reactions*. *Inputs* correspond to inputs to the signal transduction network (they are not produced by any reactions within the system). *Outputs* correspond to outputs of the signal transduction network. Unfortunately, because there is no complete knowledgebase of every reaction that occurs within the cell, some inputs may be modified phosphoproteins or complexes; not all outputs are degradation products or transcription factors. *Reactions* will attempt to maximise (or minimise, respectively) total reaction activity within the cell given your specified activities.

5) Select output type. *Full Cytoscape network* produces a network representation of the entire Reactome system. This requires visual tweaks in Cytoscape to become useful – otherwise, it will appear as a 'hairball' structure. *Differential Cytoscape network (individual comparison networks, attribute files)* will produce a differential network output – each pairwise comparison will generate a graph showing only where the two model states differ. *Differential Cytoscape network (union of comparison networks, attribute files)* produces a merged differential network showing the combination of all differences across all pairwise comparisons as a single file.

6) Select output directory, and add an experiment descriptor.

7) Specify whether you are saving mid-confidence predictions/low-confidence predictions. Mid-confidence predictions include the region of reaction system up to three reactions downstream of the difference network. Low-confidence predictions include all reactions downstream of the difference network. These should be interpreted as being potentially disrupted points in the network – they are not directly affected (and thus in the high-confidence prediction set) due to buffering, but may be affected due to upstream changes in the system.

8) Click 'OK' – simulation may take some time and require considerable computational resources.

Importing data to Cytoscape

NOTE: FastDAS' output is valid input for Cytoscape 3.2.*. It may load into Cytoscape 2.8, but there are some differences between how these two versions handle input files, particularly around node names. If you run into problems with your network loading into Cytoscape, a good place to start checking for issues are special characters used in display names and RDFs – check for characters like { } [ ] , + / \ and replace where possible.

1) From Cytoscape's 'File' menu, select 'Import > Network > File'. Select the differential network analysis file (with the .gml extension).
2) From Cytoscape's 'File' menu, select 'Import > Table > File'. Select the node attribute file (which will have a default name in the format 'NodeAttribute<timestamp>.txt'.
3) In the 'Import Columns from Table' dialog box, check the 'Show text file import options' box. In the options that appear, uncheck 'Transfer first line as column names'. Make sure that 'tab' is selected as the delimiter, then click OK.
4) Modify/arrange/layout the network within Cytoscape to suit your requirements.

Replication of analyses

An archive containing the files (and outputs) for the simulations described in the manuscript describing the FastDAS method are supplied in examples.zip.

To simulate EGF stimulation: Use the model file /Data – Curated and Original/curatedModel.owl. The pathway exclusion file is /EGF/EGF_sim_pathways.txt, and the species file is /EGF/EGF_species.txt. Select '*Maximize*' as the objective direction, '*Reactions*' as the objective, and '*Differential Cytoscape network (individual comparison networks, attribute files)*' as the output type. Check both confidence levels for output.

To simulate IL-2 stimulation: Use the model file /Data – Curated and Original/curatedModel.owl. The pathway exclusion file is /IL-2/IL2_pathways.txt.

Species files vary – for IL2 stimulation by itself, select /IL-2/IL2_only_species.txt. For IL2 with IL6 stimulus, select /IL-2/IL2_IL6_species.txt. For the experiment adding p21/RAS stimulation downstream of IL2 (patching the missing link in the Reactome version used), select /IL-2/p21_RAS_species.txt.

Select '*Maximize*' as the objective direction, '*Reactions*' as the objective, and '*Differential Cytoscape network (individual comparison networks, attribute files)*' as the output type. Check both confidence levels for output.

Troubleshooting

Most problems with FastDAS appear to be linked to the installation of the glpk libraries (all platforms) or the availability of a suitable window manager (some variants of Linux).

If all windows appear in FastDAS, but no output is produced, a log will be generated and stored in the FastDAS folder. At the time of writing, 95% of these issues are

related to the installation of glpk and glpk-java. The package documentation for these libraries provides details for installing and testing the installation of these pieces of software – this is the best start point for troubleshooting.

If FastDAS does not open a window at all, make sure you have a current installation of Sun Java and try opening FastDAS using the command prompt options. You *should* be able to run FastDAS under the OpenJDK, however this is not currently tested. If you are running Linux, problems seem to result from the use of non-tiling window managers. Unfortunately the wide number of window managers available prevents us from offering support/troubleshooting suggestions for this problem – users are recommended to seek out instructions for getting Java Swing applications running under your window manager of choice.