TENET (Target Characterization using Network Topology)

USER GUIDE (Version 1.0)

Chua Huey Eng

10/20/2014

Contents

1. TENET Installation

TENET uses several other publicly available tools and software for pre-processing files and storing data. In particular, the PostgreSQL database management system is used for storing data. PostgreSQL has to be installed in order for TENET to function. Please adhere to the recommended version as it has been tested. Note that TENET has been tested for installation on the Windows 7 Professional platform.

1.1 PostgreSQL 9.3

Installing PostgreSQL 9.3
Step 1: Download and install PostgreSQL v9.3.x from http://www.postgresql.org/.

Please refer to http://www.postgresqltutorial.com/install-postgresql/ for additional guidance on installation using the PostgreSQL installer for Windows.

Step 2: During the setup progress, when prompted for password (Fig.1), type in “tenet”.

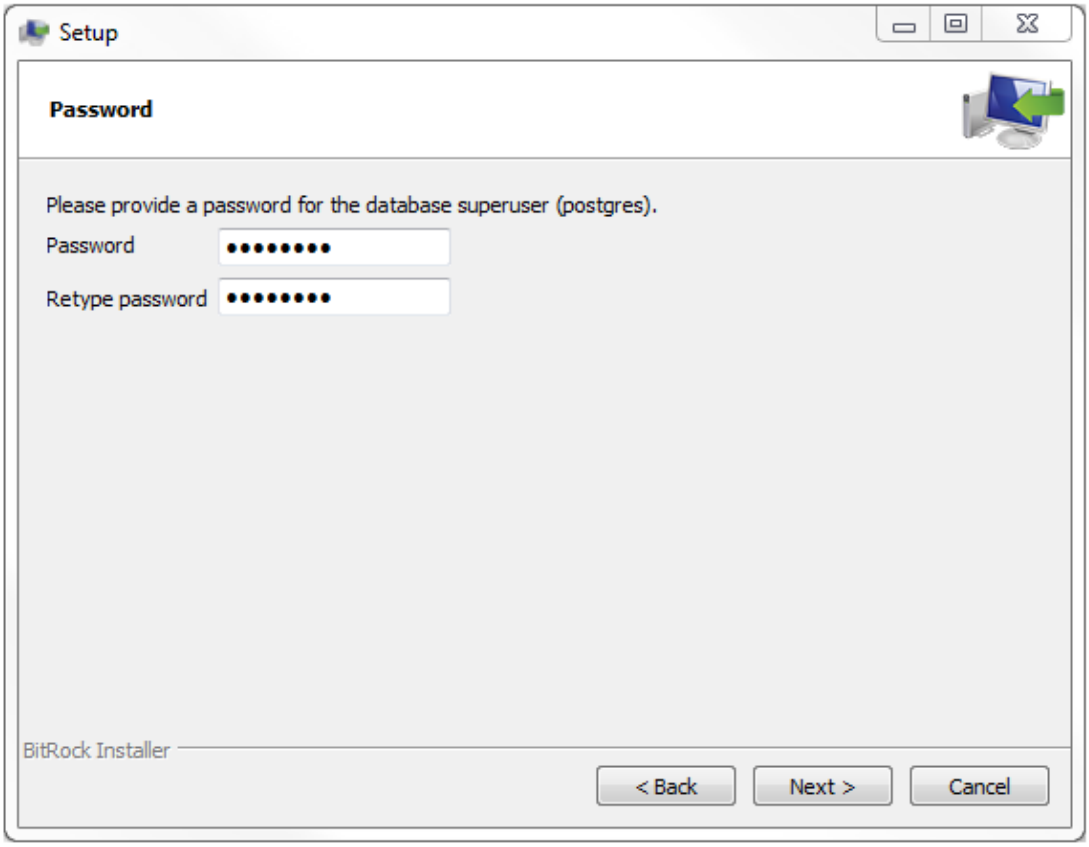


Fig. 1: Password prompt during PostgreSQL installation on Windows.

Verifying the installation
Step 3: Select “pgAdmin III” from programs to launch PostgreSQL (Fig. 2)

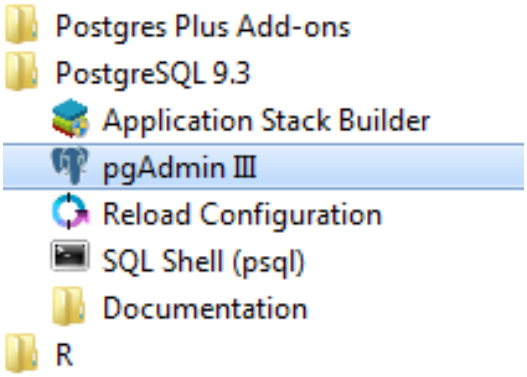


Fig. 2: Launching PostgreSQL

Step 4: Double click on PostgreSQL 9.3 on the object browser and enter “tenet” as the password when prompted (Fig. 3).

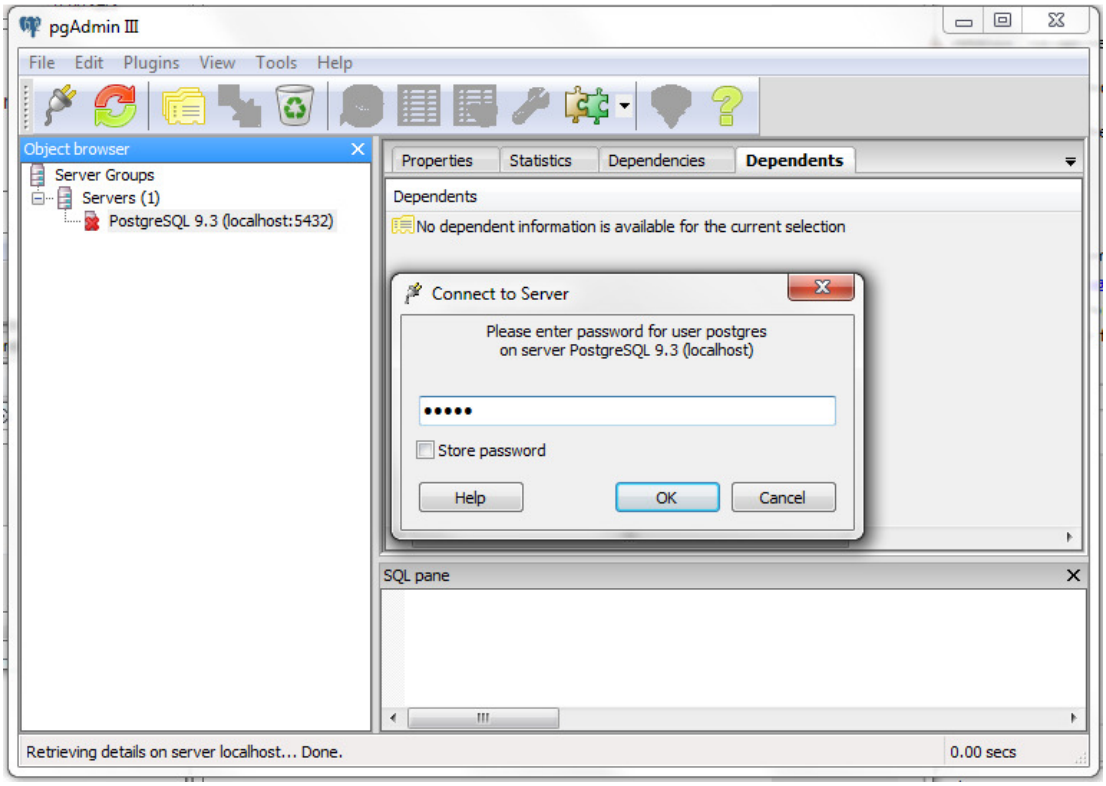


Fig. 3: Connecting to PostgreSQL server

Creating the TNKB database

The TNKB database is used by TENET for storing data and has to be created before running TENET. The database creation has to be done only once during the installation phase and will be available subsequently when TENET is run.

Step 5: Right-click on “Databases” on the object browser and select “New Database…” (Fig. 4)

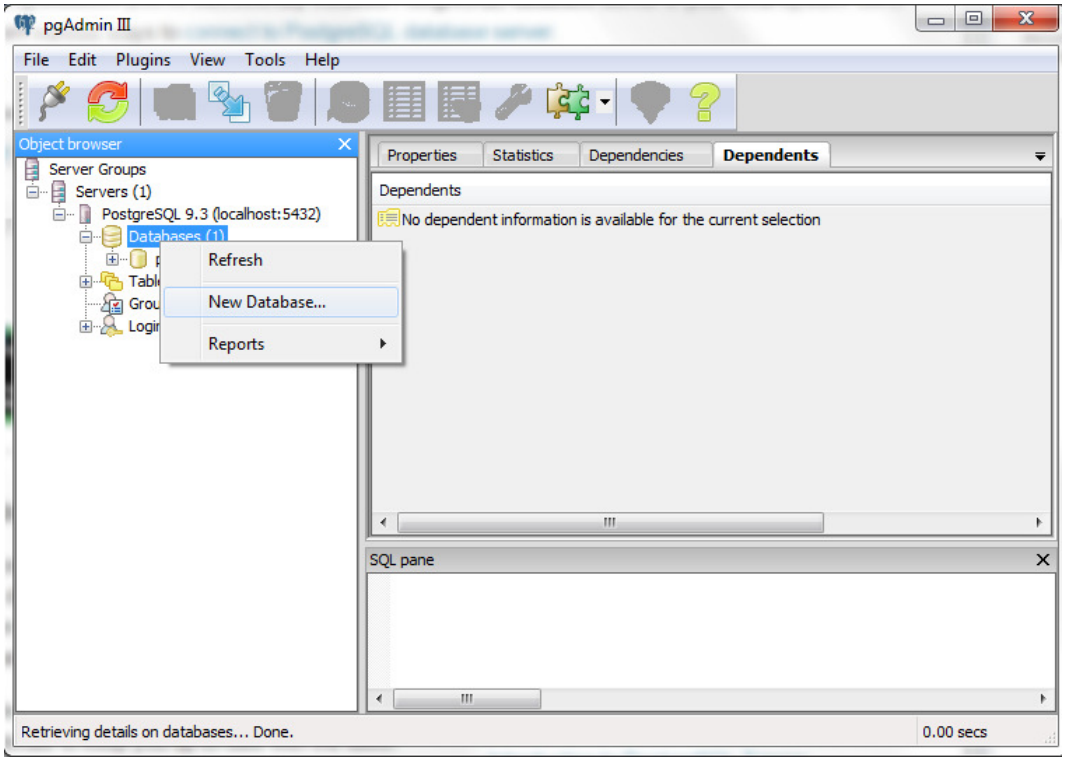


Fig. 4: Creating new database

Step 6: Type in “TNKB” as the name of the database (Fig. 5) and select “OK”.

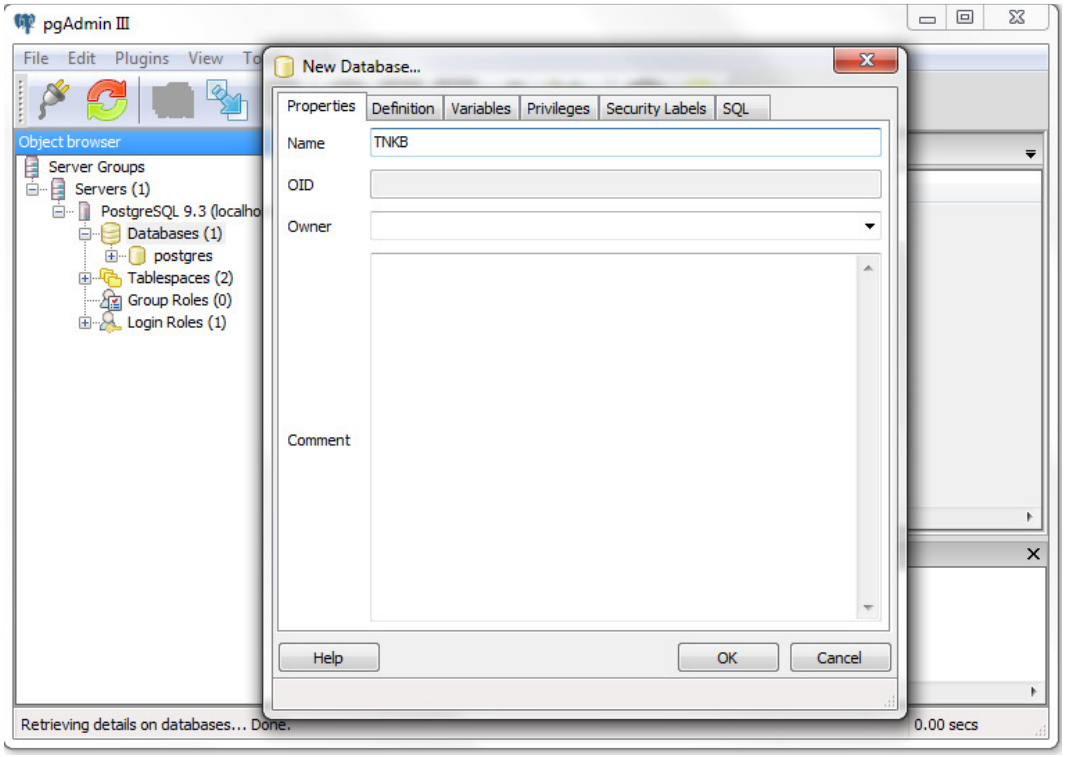


Fig. 5: Setting the name of the new database

Step 7: Verify that the “TNKB” database has been created

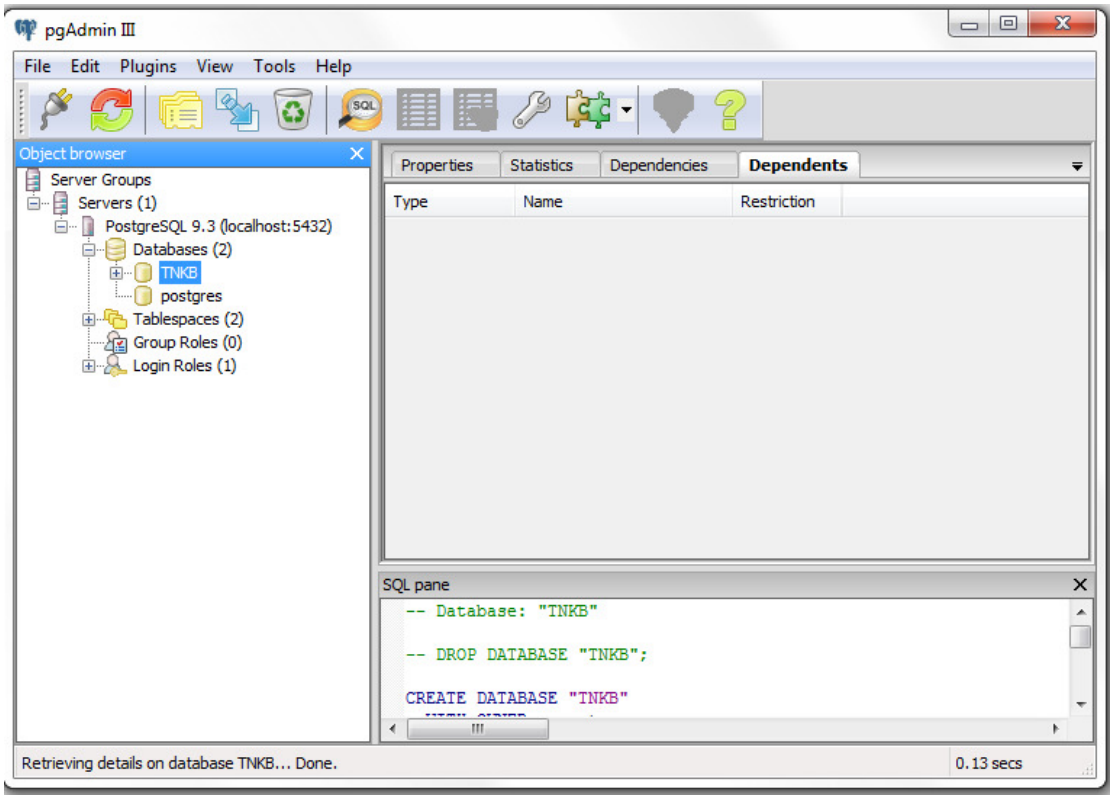


Fig. 6: Verify that the new database has been created

1.2 network_lib folder

This folder contains the example files for running TENET. TENET uses SVM to perform target prioritization and intermediate files generated while running the SVM are stored in the “svmTraining” folder within the “network_lib” folder.

TENET requires several files from this folder for proper execution:

1. lib_list.txt
   – contains the list of files (e.g., hatakeyama2003_MAPK_biomodels.xlsx) containing the known target information of the example signalling networks that can be selected by TENET for target characterization.

2. <signalingNetworkTargetInformation>.xlsx
   – contains the target information of the signalling network. An example is hatakeyama2003_MAPK_biomodels.xlsx. A description of this .xlsx format is provided later in this document for users interested in modifying the target information of the example networks or to add new example networks for analysis in TENET. During execution, TENET generates three intermediate files “network.csv”, “source.csv” and “target.csv”.

3. <signalingNetwork>.xml

– contains the signalling network model that is described in SBML format. An example is Hatakeyama2003_MAPK.xml. The BioModels repository (http://www.ebi.ac.uk/biomodels-main/) provides signalling networks that are described in the SBML format.

Step 8: Copy the “network_lib” folder to the C directory “C:\” (Fig. 7)

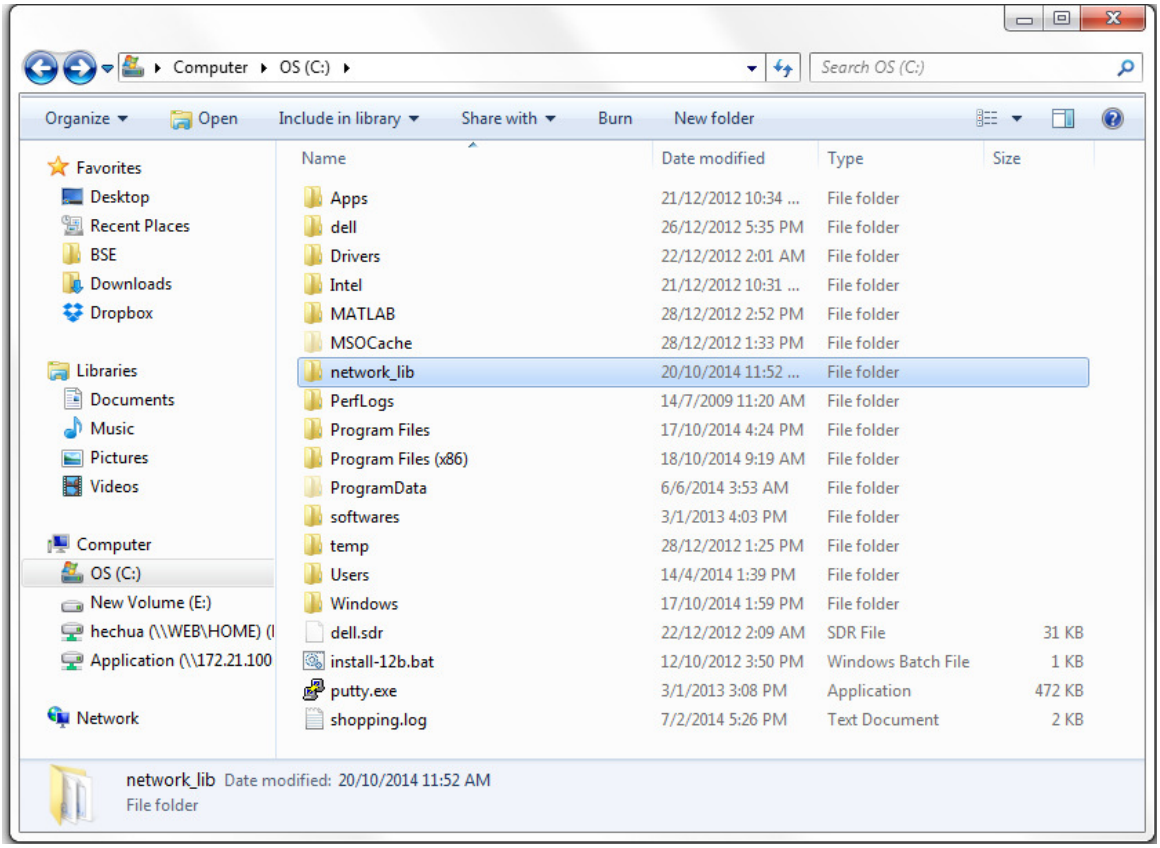


Fig. 7: Copy “network_lib” folder to C directory

1.3 TENET.jar

Step 9: Copy TENET.jar to a desired location (e.g., “C:\Desktop”).

2. Launching TENET

Step 1: Launch command prompt dialog (cmd.exe) (Fig. 8).

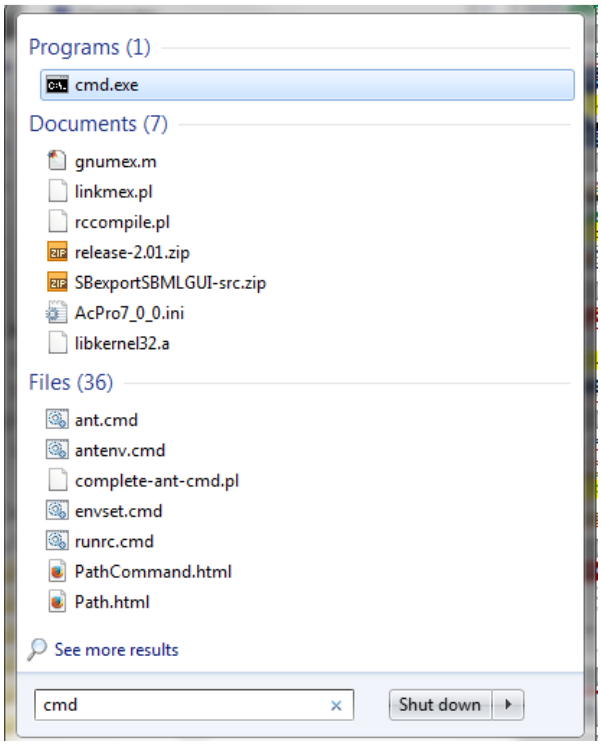


Fig. 8: Launch cmd.exe

Step 2: Change the directory to the one that contains TENET.jar (Fig. 9).

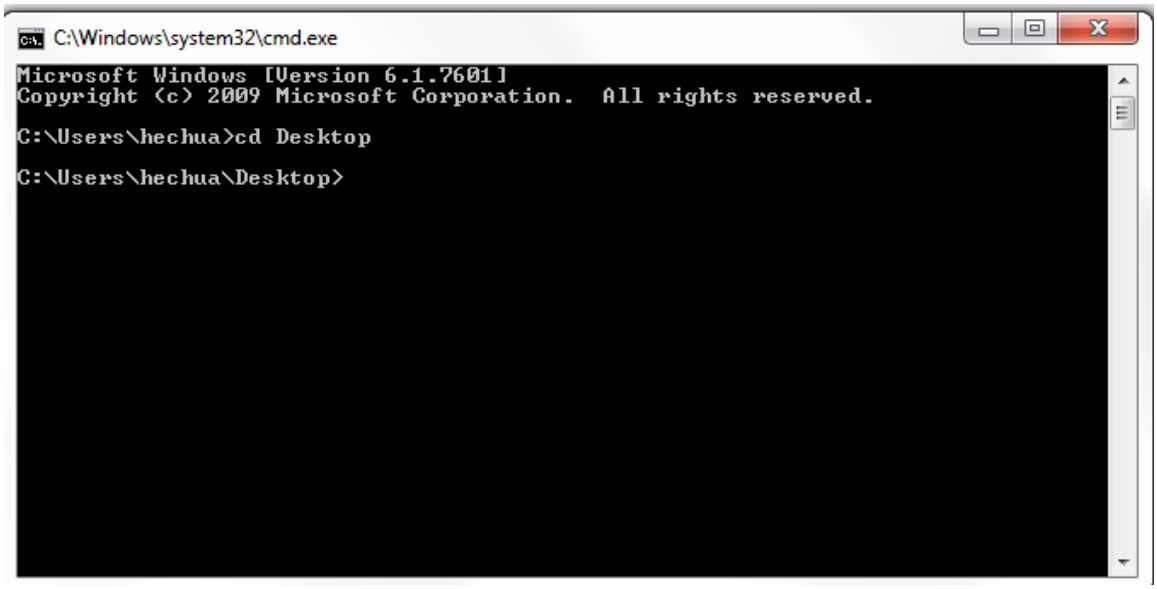


Fig. 9: Set to the directory containing TENET.jar

Step 3: Run TENET by using the command “java -jar TENET.jar” (Fig. 10).

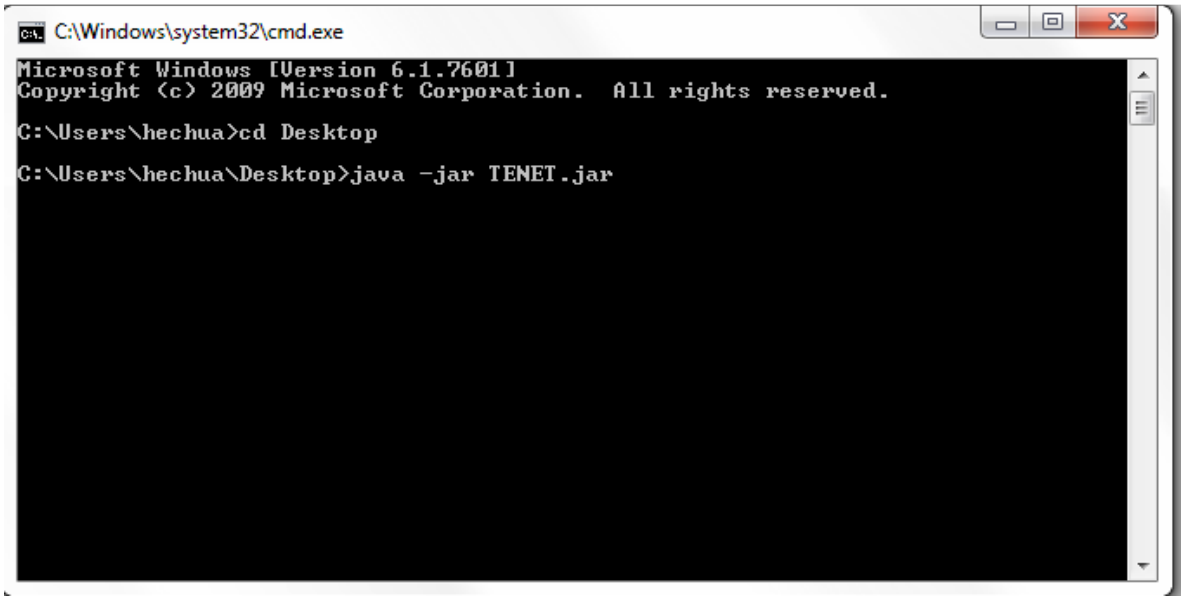


Fig. 10: Run TENET

3. Using TENET

Step 1: Select the signalling network to be characterized (Fig. 11).

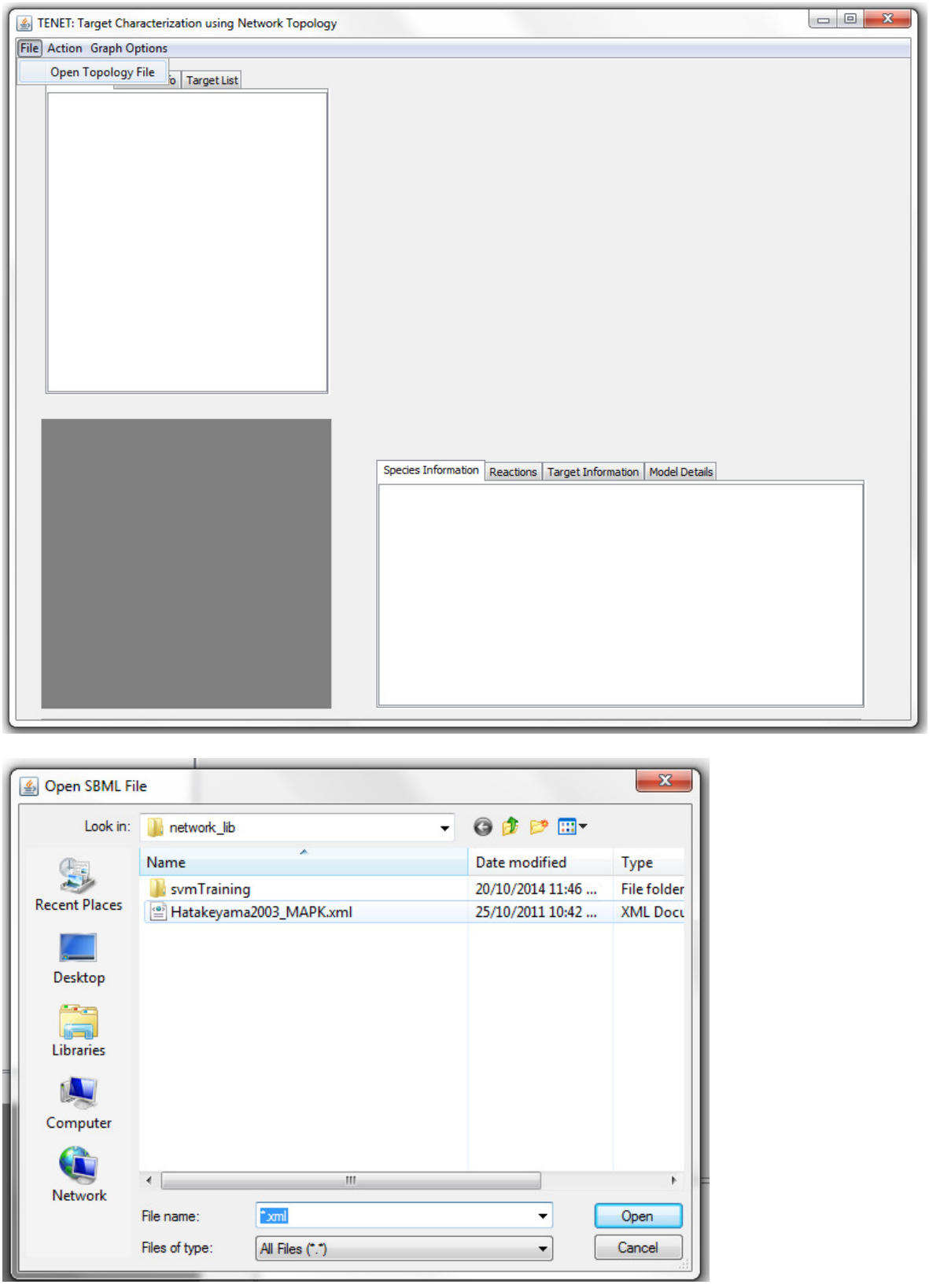


Fig. 11: Select signalling network

Step 2: Select the appropriate organism and disease for the input signalling network (Fig. 12).

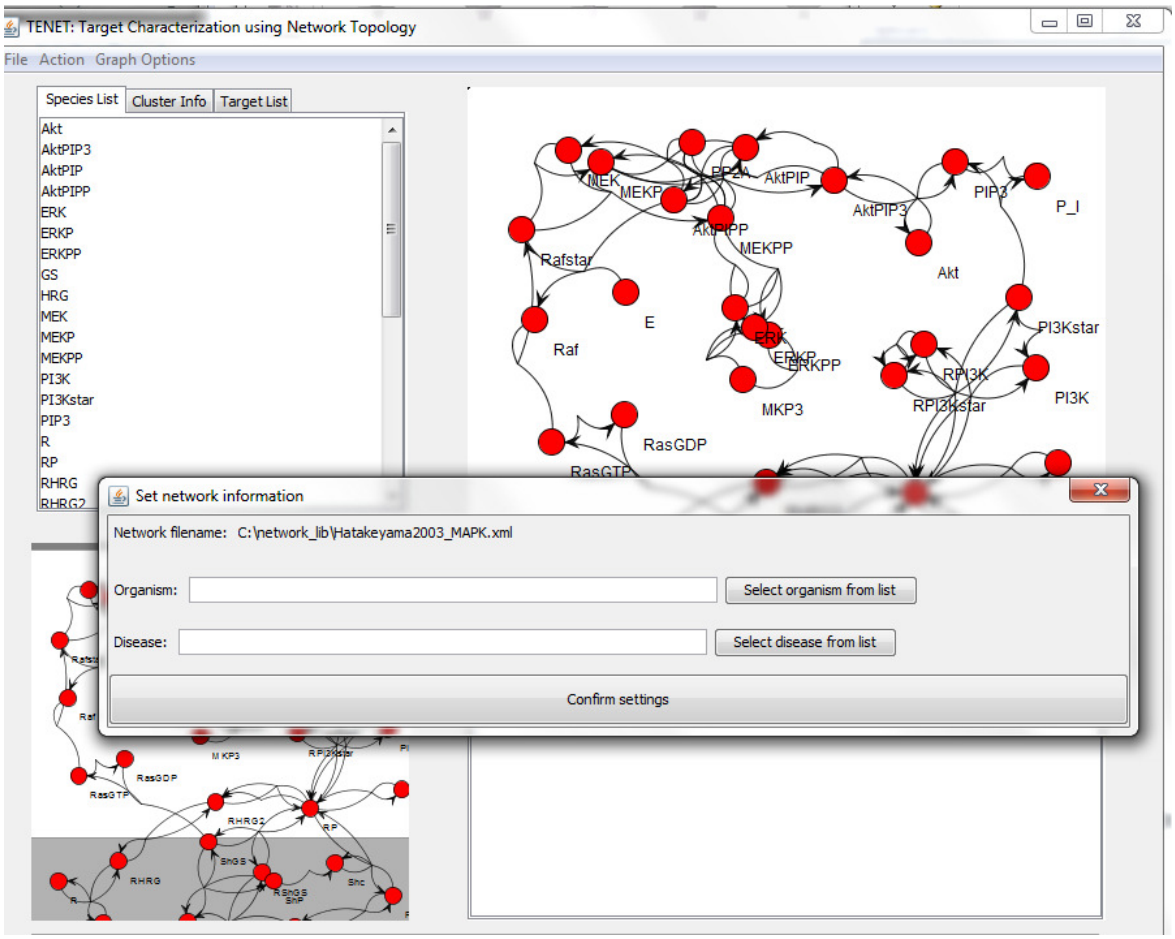


Fig. 12: Select appropriate organism and disease for the signalling network

3.1 Viewing target information

The “Target List” provides a list of targets for this network (Fig. 13). Selecting a target from the list displays the target information in the “Target Information” panel. The information listed in this panel includes the drug and target information and the clinical trials related to the drugs. URL links are provided for the information and user can click directly on these links to get access to the information online.

3.2 Characterizing targets

TENET pre-processes the network before characterizing the targets. One of the steps in the pre-processing is to prune irrelevant nodes and it does so by checking the reachability of a specific node (output node) in the network that has direct link to the disease or biological phenomenon the network is related to. This output node is chosen by the user usually based on some biological knowledge.

Step 3: Select output node from “Species List” (Fig. 14).

Step 4: Launch the target characterization dialog by selecting “Action” and then “Characterize Target” (Fig. 15).

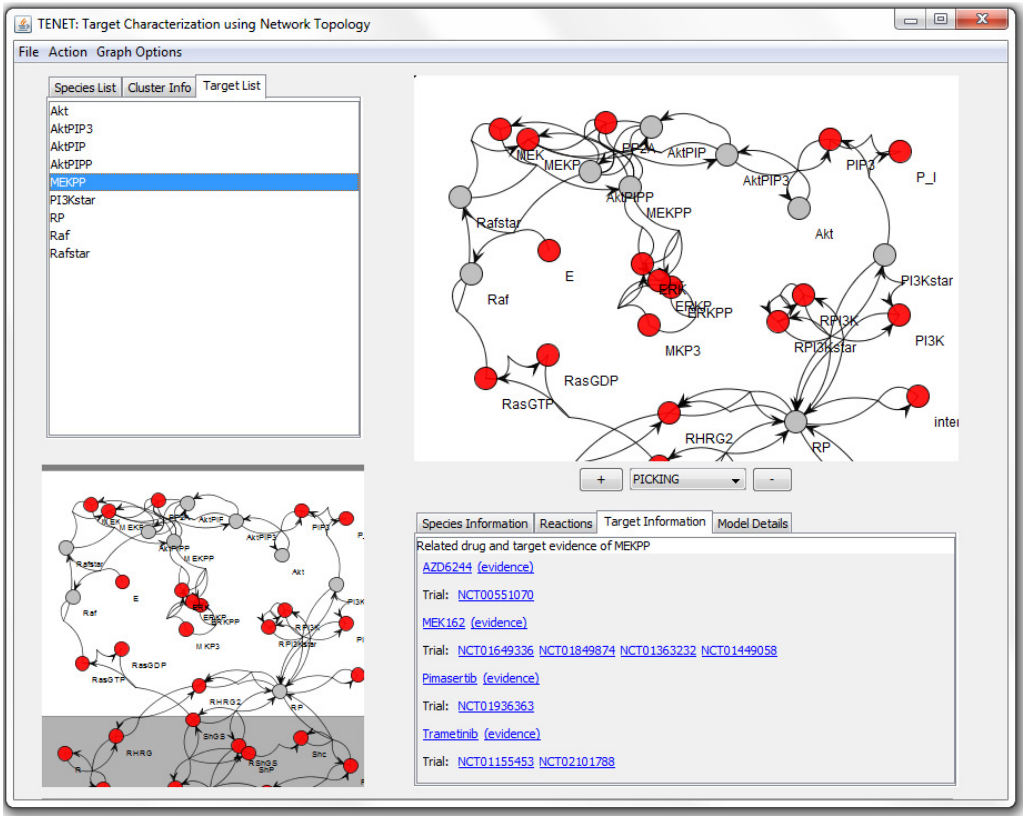


Fig. 13: Viewing target information

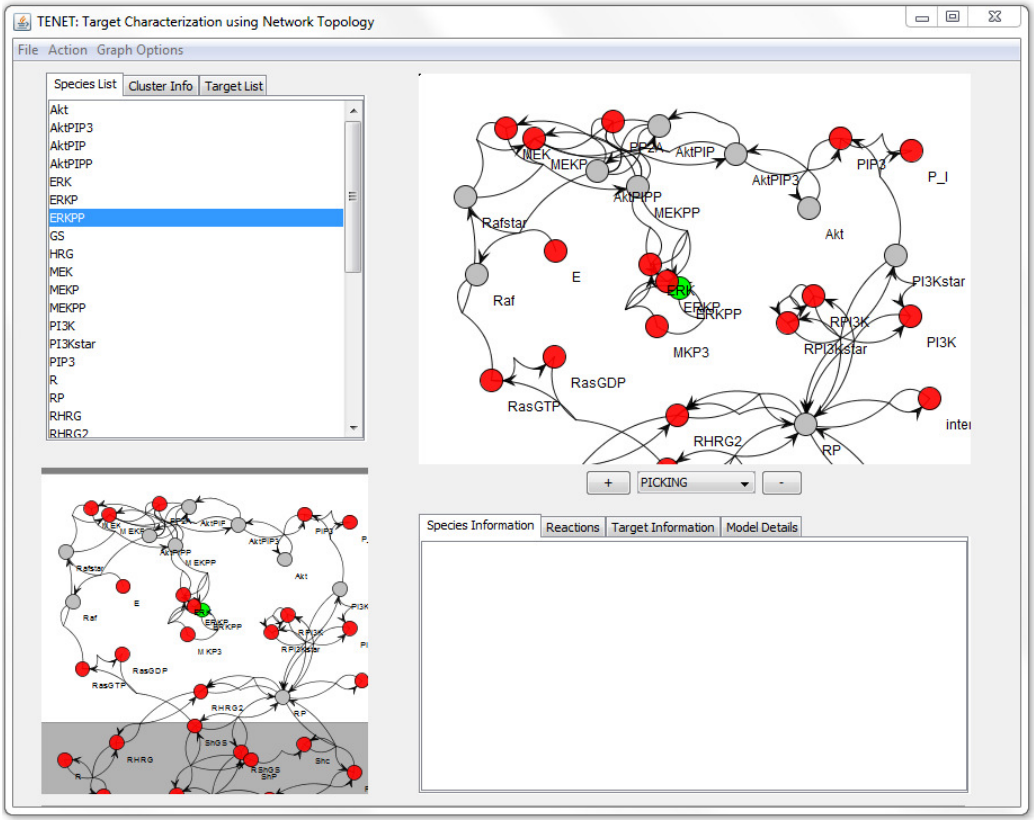


Fig. 14: Select output node “ERKPP” from “Species List”.

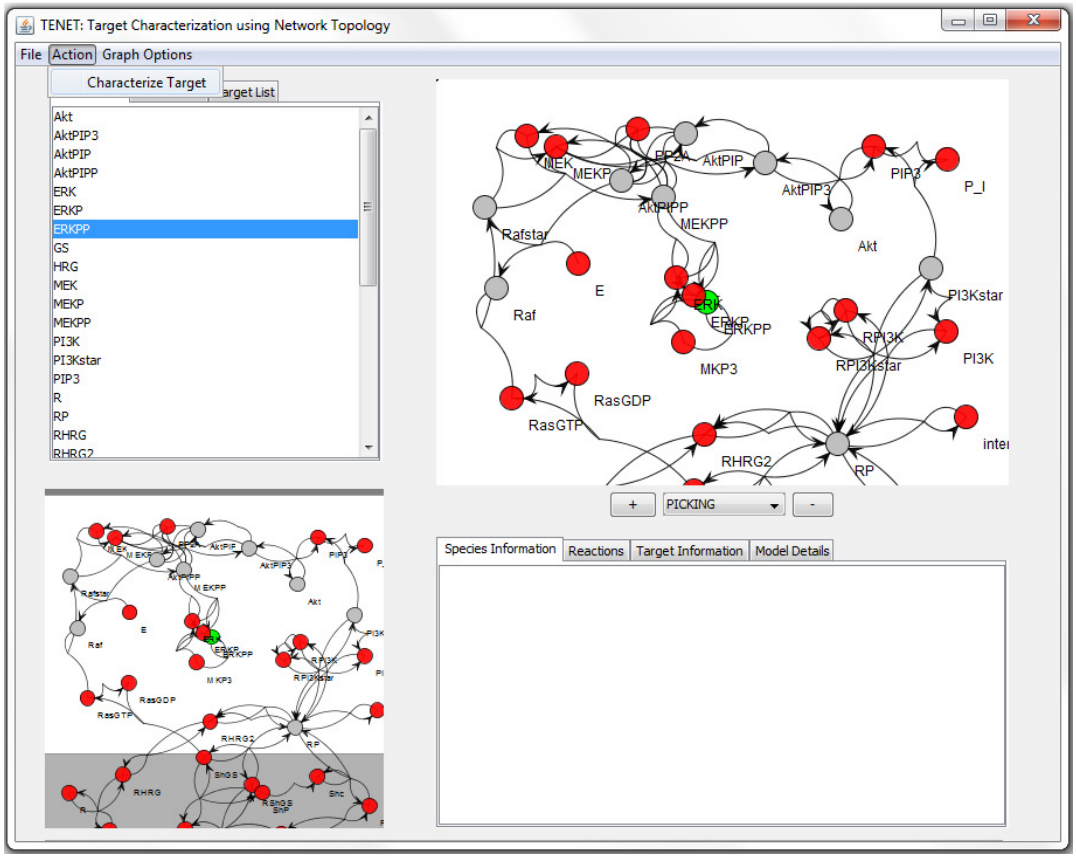


Fig. 15: Launch the dialog for target characterization.

Step 5: Configure the settings for target characterization and perform target characterization by clicking on “Characterize targets” (Fig. 16).

The configurable settings include:

- List of targets
- List of features to be used for characterization
- Number of cross-validation folds to use
- Feature selection approach
- SVM kernel
- Weighted misclassification cost option

By defaults,

- The list of targets include all targets in the <signalingNetworkTargetInformation>.xlsx (refer Section 1.2).
- The list of features include all the 16 features listed
- Number of cross-validation is set to 10 (when there are at least 10 targets) or <numOfTargets>-1 (when there are less than 10 targets). The latter is to ensure that there is at least one target in the test set

- Feature selection approach is set to NONE (naïve SVM). Three other feature selection approaches, namely, backward stepwise elimination (BSE), Wilcoxon-ROC elimination (WRE) and a hybrid approach involving WRE and BSE, are provided.
- SVM kernel is set to linear. We currently only support linear kernel.
- Weighted misclassification cost option is not selected. Selecting the weighted misclassification cost will result in TENET assigning a SVM cost variable for misclassifying the target to be different from that for misclassifying the non-target. A range of misclassification cost will be tested and the one resulting in the best prediction accuracy for the cross-validation shall be chosen for the final SVM model.

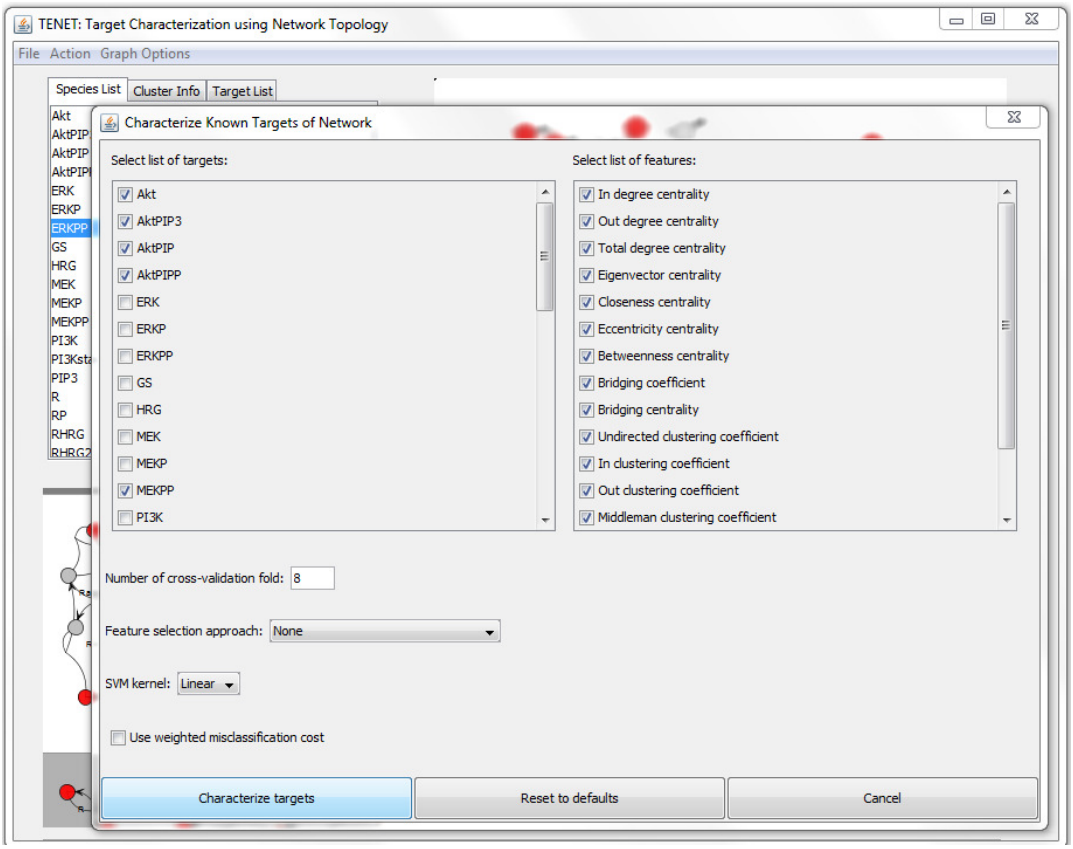


Fig. 16: Settings for target characterization.

The details of the SVM model are provided in “Model Details” (Fig. 17). The details include:

- Type of kernel
- SVM cost parameter
- Target misclassification cost. Note that a value of 0 indicates that the weighted misclassification cost option is not selected
- Validation accuracy. This is the prediction accuracy of the cross-validation
- Test accuracy. This is the prediction accuracy of the test set
- List of predictive structural features used for constructing the final SVM

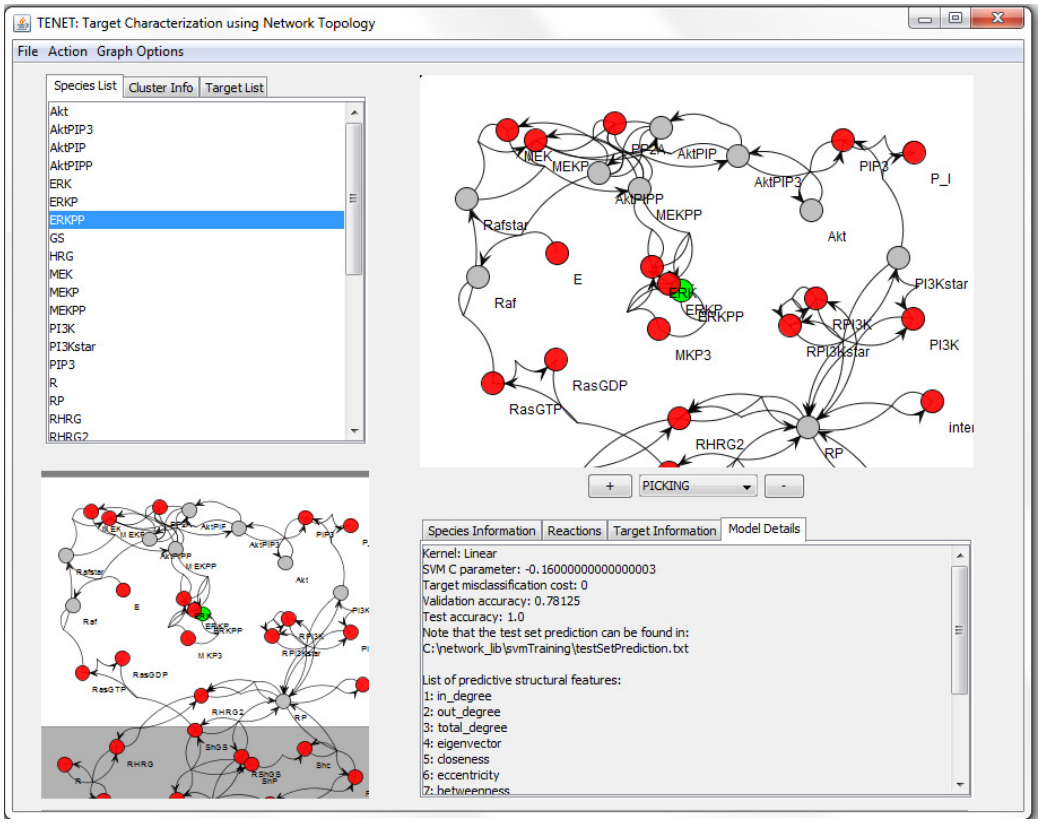


Fig. 17: Model details of the final SVM model.

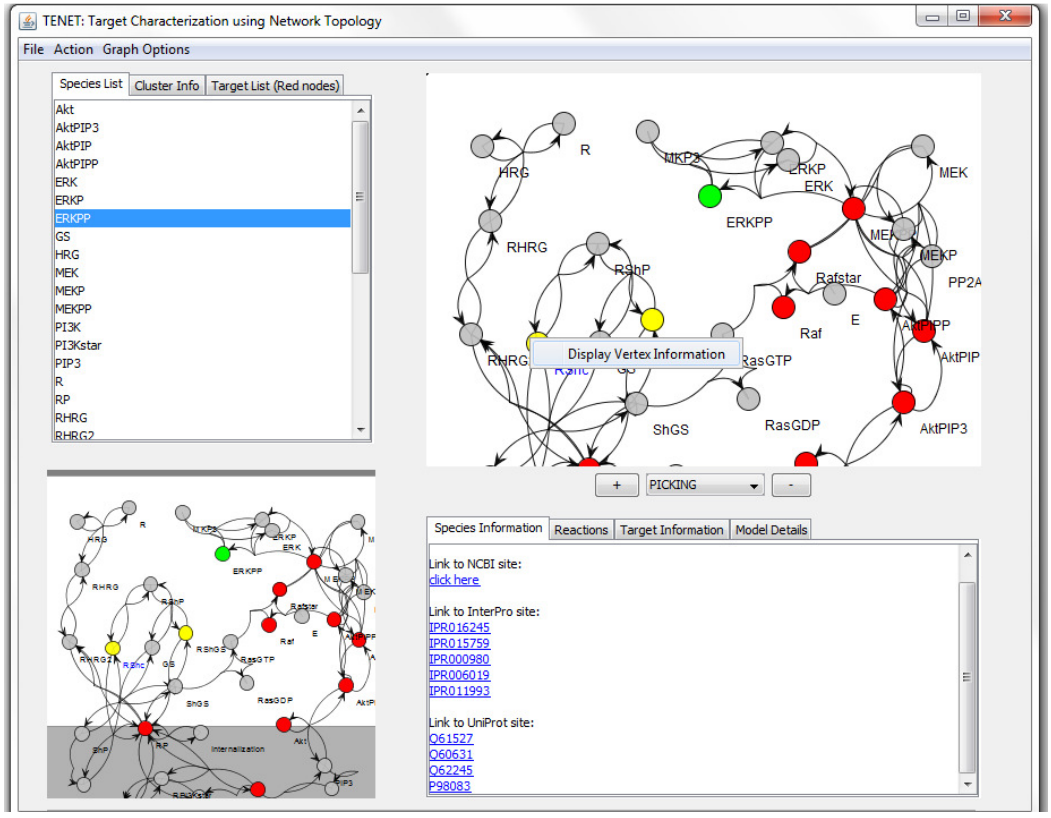


Fig. 18: Viewing node information.

3.3 Viewing node information

The information related to the node can be viewed in the “Species Information” panel by right-clicking the node in the interactive graph panel (top right panel in Fig. 18) and selecting “Display Vertex Information” in the pop-up menu.

4. Adding new signalling network for characterization

Step 1: Create the signalling network target file <signalingNetworkTargetInformation>.xlsx for the new signalling network and save it in the network_lib folder.

Users may use the signalingNetworkTarget_template.xlsx in the network_lib folder for creating this file. An example of a signalling network target file is hatakeyama2003_MAPK_biomodels.xlsx.

The file contains three tabs (“target”, “source”, “network”):

- “target” tab – Contains information about the target.
- “source” tab – Contains information about the data source.
- “network” tab - Contains information about this signalling network.

4.1 “target” tab

In the “target” tab, the rows represent the targets and the columns represent the properties of the targets (Fig. 18).

	A	B	C	D	E	F	G	H	I	J	
1	entry	target_name	target_source	target_source_id	drug_name	drug_source	drug_source_id	evidence_source	evidence_source_id	trial_source	trial_source_id
2	1	RP	Others		Lapatinib	DrugBank	DB01259	PubMed	18334220	ClinicalTrials.gov	NCT00317434 NCT00888810 NCT00316407 N
3	2	Rafstar	Others		Sorafenib	DrugBank	DB00398	PubMed	17016424	ClinicalTrials.gov	NCT00526799 NCT00096395 NCT01047891 N
4	3	Rafstar	Others		Dabrafenib	DrugBank	DB08912	NCIThesaurus	C82386	ClinicalTrials.gov	NCT01902173
5	4	Raf	Others		ISIS 5132	NCIThesaurus	C2721	PubMed	10815879	ClinicalTrials.gov	NCT00003892
6	5	Raf	Others		ECO-4601	DrugBank	DB05923	NCIThesaurus	C62508	ClinicalTrials.gov	NCT00338026
7	6	MEKPP	Others		AZD6244	NCIThesaurus	C66939	PubMed	17332304	ClinicalTrials.gov	NCT00551070
8	7	MEKPP	Others		MEK162	NCIThesaurus	C84865	NCIThesaurus	C84865	ClinicalTrials.gov	NCT01649336 NCT01849874 NCT01363232 N
9	8	MEKPP	Others		Pimasertib	NCIThesaurus	C84864	NCIThesaurus	C84864	ClinicalTrials.gov	NCT01936363
10	9	MEKPP	Others		Trametinib	DrugBank	DB08911	PubMed	21245089	ClinicalTrials.gov	NCT01155453 NCT02101788
11	10	PI3Kstar	Others		XL147	DrugBank	DB05240	NCIThesaurus	C71705	ClinicalTrials.gov	NCT00756847
12	11	PI3Kstar	Others		PKI-587	NCIThesaurus	C91732	PubMed	20166697	ClinicalTrials.gov	NCT00940498 NCT02069158
13	12	PI3Kstar	Others		PKI-179	NCIThesaurus	C90291	NCIThesaurus	C90291	ClinicalTrials.gov	NCT00997360
14	13	PI3Kstar	Others		BKM120	NCIThesaurus	C90565	NCIThesaurus	C90565	ClinicalTrials.gov	NCT01068483 NCT01623349 NCT01155453 N
15	14	PI3Kstar	Others		BYL719	NCIThesaurus	C94214	NCIThesaurus	C94214	ClinicalTrials.gov	NCT01708161 NCT01449058
16	15	PI3Kstar	Others		SAR245409	NCIThesaurus	C71704	NCIThesaurus	C71704	ClinicalTrials.gov	NCT01936363
17	16	AktPIP	Others		Perifosine	NCIThesaurus	C1727	PubMed	14617782	ClinicalTrials.gov	NCT00431054
18	17	AktPIPP	Others		Perifosine	NCIThesaurus	C1727	PubMed	14617782	ClinicalTrials.gov	NCT00431054
19	18	AktPIP3	Others		Perifosine	NCIThesaurus	C1727	PubMed	14617782	ClinicalTrials.gov	NCT00431054
20	19	Akt	Others		AZD5363	NCIThesaurus	C102564	NCIThesaurus	C102564	ClinicalTrials.gov	NCT01226316

target source network

Fig. 18: “target” tab in hatakeyama2003_MAPK_biomodels.xlsx

Column headers:

- entry
  – This counts the number of entries in the tab. The count starts from 1.
- target_name
  – This is name of the target and must correspond to the ID of the node in the network described in SBML format.
- target_source
  – This describes the source which contains information of the node. Examples of possible sources are UniProt databases (“UniProt”) or online publications (“Others”).
- target_source_id

– This provides details of how the target information shall be retrieved from the *target_source*. For example, the URL of the online publication.

- drug_name
  – This is the name of the drug that targets the node.
- drug_source
  – This describes the source which contains information regarding the drug.
- drug_source_id
  – This provides details of how the drug information shall be retrieved from the *drug_source*. For example, the drug ID of the drug in the *drug_source*.
- evidence_source
  – This describes the source which contains evidence that the node is a target or that a drug targets a particular node.
- evidence_source_id
  – This provides details of how the evidence information shall be retrieved from the *evidence_source*.
- trial_source
  – This describes the source which contains trial information that the drug is in clinical trial for a particular disease
- trial_source_id
  – This provides details of how the trial information shall be retrieved from the *trial_source*.

4.2 “source” tab

In the “source” tab, the rows represent the data source and the columns represent the properties of the data source (Fig. 19).

	A	B
1	source_name	source_url
2	PubMed	www.ncbi.nlm.nih.gov/pubmed/<v>
3	DrugBank	www.drugbank.ca/drugs/<v>
4	NCIThesaurus	http://ncit.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=N
5	Blood	www.bloodjournal.org/content/<v>
6	PMC	www.ncbi.nlm.nih.gov/pmc/articles/<v>
7	ClinicalTrials.gov	https://clinicaltrials.gov/ct2/show/<v>
8	BioModels	http://www.ebi.ac.uk/biomodels-main/<v>
9	KEGG disease	http://www.genome.jp/dbget-bin/www_bget?ds:<v>
10	KEGG genome	http://www.genome.jp/dbget-bin/www_bget?gn:<v>
11	Others	<v>
12		

target | source | network

Fig. 19: “source” tab in hatakeyama2003_MAPK_biomodels.xlsx

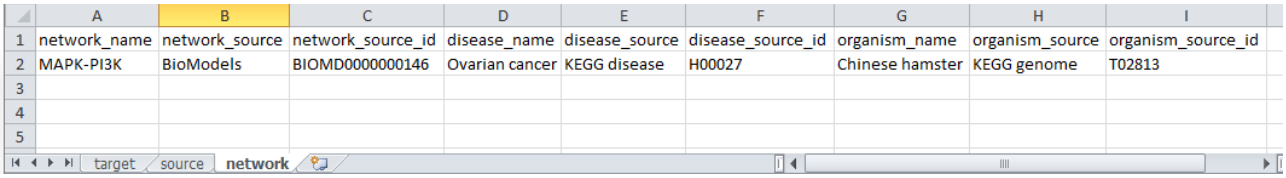
Column headers:

- source_name
  – The source name that is referenced by *target_source*, *drug_source*, *evidence_source* and *trial_source* in “target” tab.

- source_url
  – This is URL format that shall be used by TENET for accessing the information online. Note that <v> shall be replaced by the source ID (*target_source_id*, *drug_source_id*, *evidence_source_id* and *trial_source_id*) in TENET.

4.3 "network" tab

In the "network" tab, the rows represent the network and the columns represent the properties of the network (Fig. 20).



	A	B	C	D	E	F	G	H	I
1	network_name	network_source	network_source_id	disease_name	disease_source	disease_source_id	organism_name	organism_source	organism_source_id
2	MAPK-PI3K	BioModels	BIOMD0000000146	Ovarian cancer	KEGG disease	H00027	Chinese hamster	KEGG genome	T02813
3									
4									
5									

target | source | network

Fig. 20: "network" tab in hatakeyama2003_MAPK_biomodels.xlsx

Column headers:

- network_name
  – The name of this signalling network.
- network_source
  – This describes the source from which this network is obtained. Examples are database repository such as BioModels.
- network_source_id
  – This provides details of how the network information shall be retrieved from the *network_source*.
- disease_name
  – The name of the disease that this network is related to.
- disease_source
  – This describes the source from which the disease information can be obtained.
- disease_source_id
  – This provides details of how the network information shall be retrieved from the *disease_source*.
- organism_name
  – The name of the organism that this network is based on.
- organism_source
  – This describes the source from which the organism information can be obtained.
- organism_source_id
  – This provides details of how the network information shall be retrieved from the *organism_source*.

Step 2: Update lib_list.txt with the signalling network target file <signalingNetworkTargetInformation>.xlsx.

Step 3: Save the <signalingNetwork>.xml file that describes the new signalling network in SBML format in the network_lib folder.