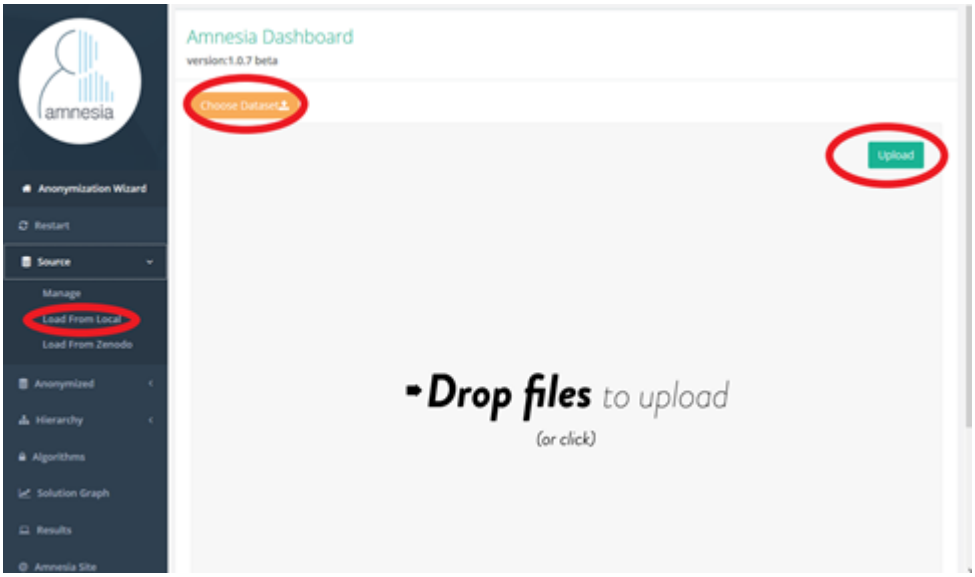
k^m-anonymity

k^m-anonymity is a weaker form of *k*-anonymity that is better suited for high-dimensional data. As in *k*-anonymity, the algorithm considers a number *n* of quasi identifiers, but now limits the guaranty against adversaries that know only *m* (any *m*) of the *n* quasi identifiers. In other works, the anonymization algorithm guarantees that each combination of *m* quasi identifiers appears *k* times in the anonymized datasets, independently of the total number *n* of quasi identifiers. Typically, $m << n$. k^m-anonymity is used for datasets with set-valued attributes in Amnesia.

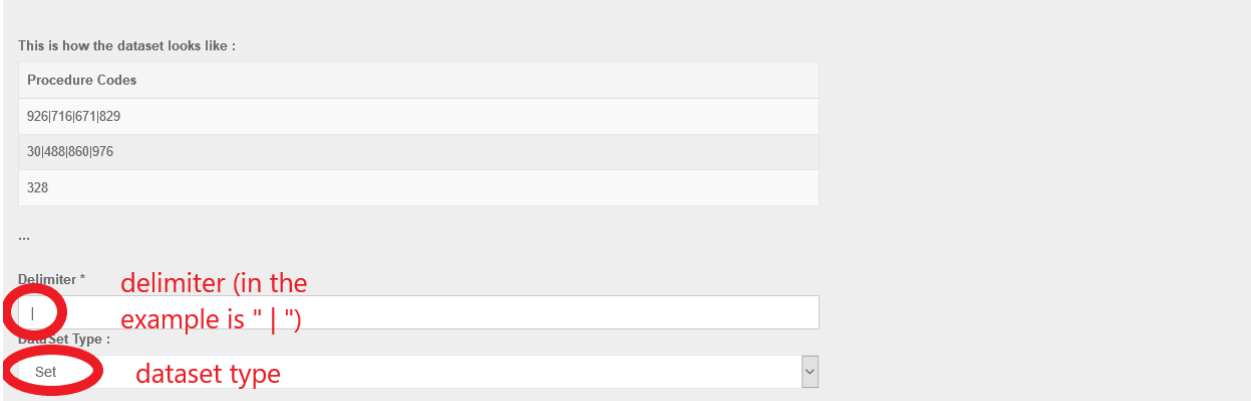
Set-values

The **Object-relational dataset** is a combination between relational dataset and set-valued dataset. An **object-relational** dataset contains an arbitrary number of columns with simple type values and a single set-valued attribute. The simple values are separated by a delimiter, different from the delimiter that indicates the columns. Datasets handled by Amnesia contain up to one set-valued attribute. The simple values can be only strings. Amnesia applies only ***k^m*-anonymity** on object relational datasets, **using** the **apriori** algorithm which produces a single solution with a locally optimum "cut" in the generalization hierarchies. The following example shows how to apply ***k^m*-anonymity** in an object relational dataset.

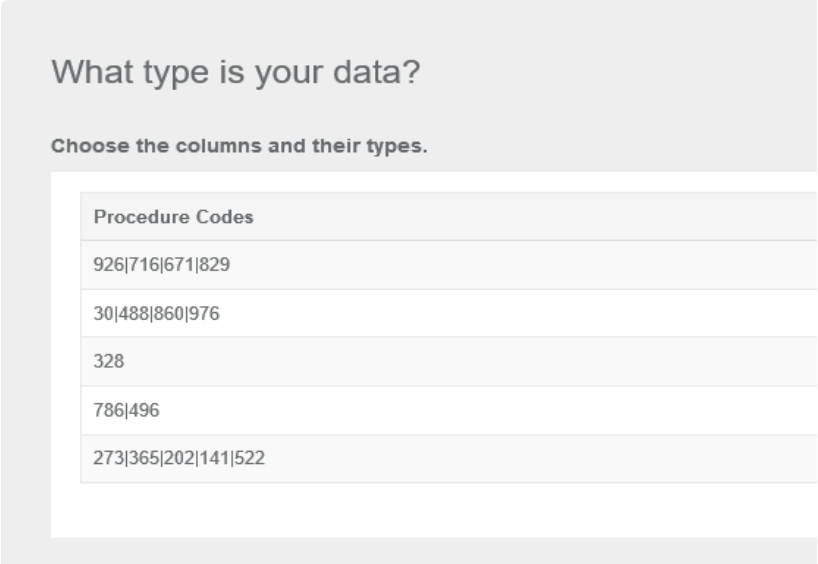
First, the user chooses the appropriate file from his/her local system.



The user must select the type of the dataset, i.e., "Relational and Set" for this example. Finally, the user must define two delimiters: a) one for the different columns and b) one for parsing the simple values inside the set-valued attribute.



Amnesia will guess the type of data, but the user has the chance to alter the estimations of Amnesia. The set-valued attribute will be of special type “set”, which implies a set of strings.

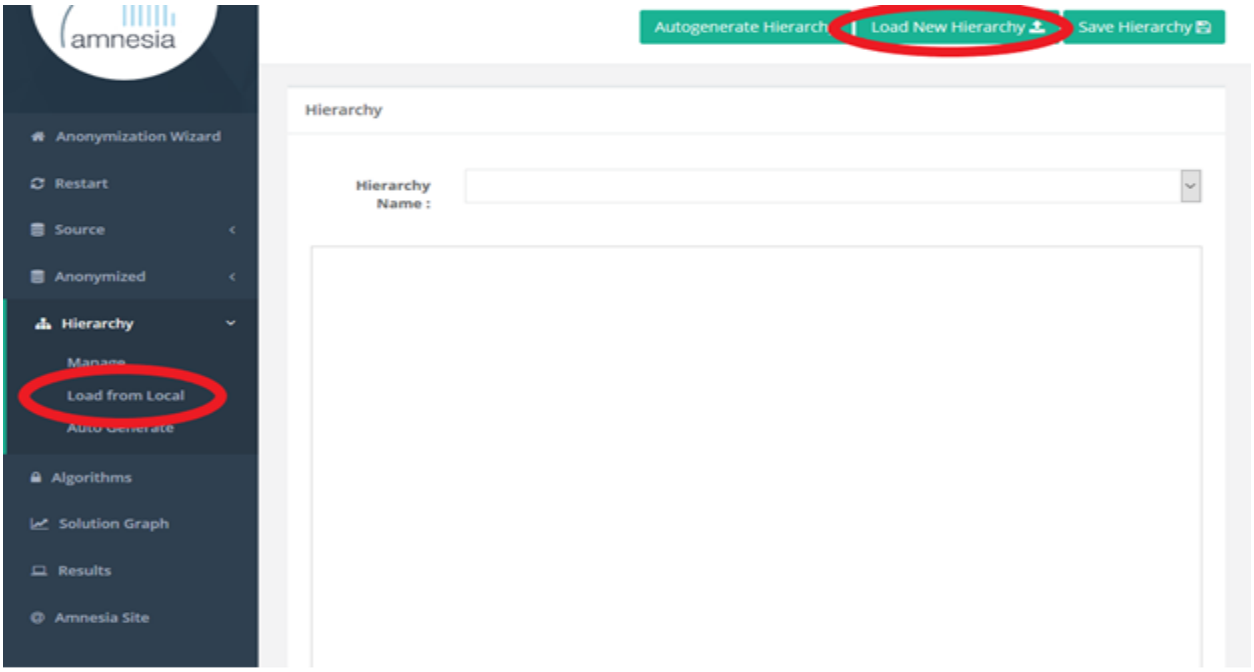


Introducing Generalization hierarchies

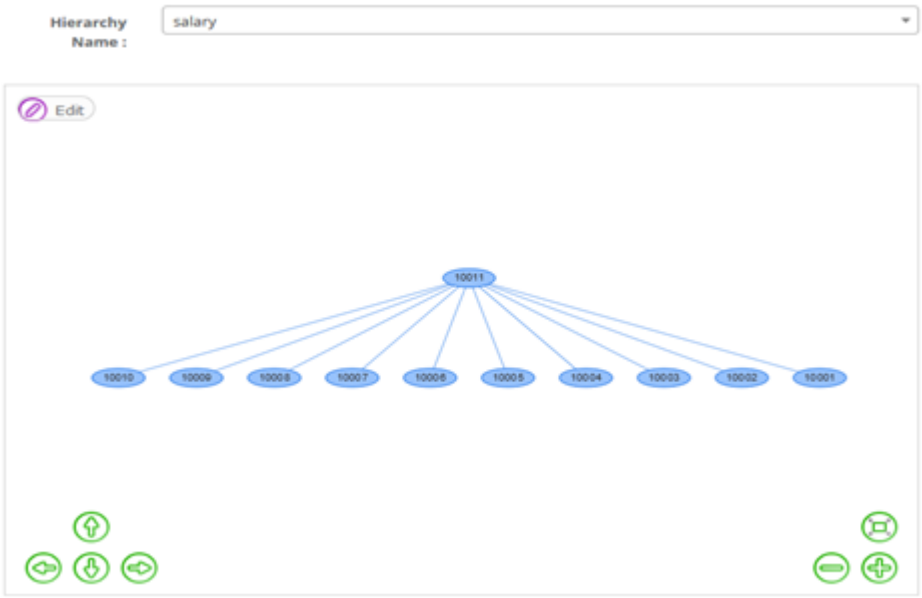
Amnesia’s k-anonymity algorithm uses generalization to create groups of identical data. Unique or rare (appearing less than k times in the dataset) values or value combinations are replaced by more generic values. For example, if there is a single person in a dataset that resides in Greece, then Greece (and the rest of EU countries) will be replaced by EU to create more than k identical values in the Country of Residence value. Generalization hierarchy that define how these replacements take place are provided by the user as input to the algorithm.

Amnesia assists the user to create hierarchies using existing datasets, more details on how to create and edit and existing hierarchy are provided in the XXXXXX Tutorial

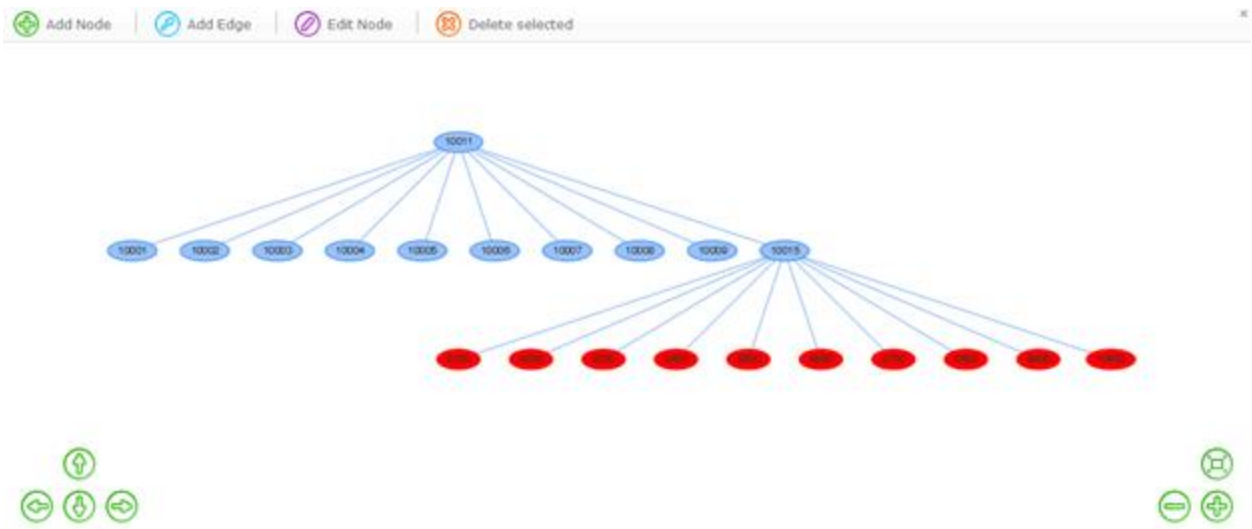
The user can upload a hierarchy in 2 different ways: a) by clicking “Load New Hierarchy” button or b) by clicking “Load from Local” in the left side menu.



Loaded hierarchies appear on the screen as graphs, where its parent node indicates how its children nodes are generalized.

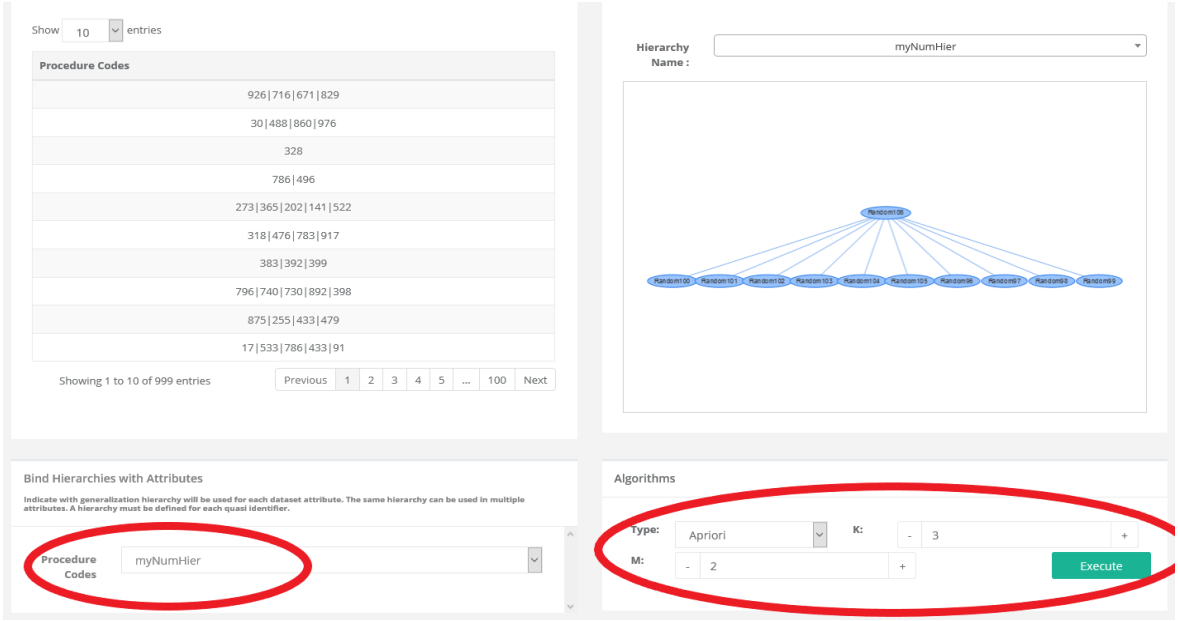


By double clicking in a node one can see its children.



Execution of Algorithms and Results

Once, the dataset and the hierarchies are loaded, the user can proceed with invoking the anonymization algorithm. The user only needs to: a) associate each attribute with the hierarchy that must be used on it and b) define the value of *k* and *m*. The execute button start the anonymization algorithm.



The algorithm is ready to be executed. After, the execution the anonymized and the original dataset side by side are presented in the screen. The user can download neither the anonymized dataset or the anonymization rules.