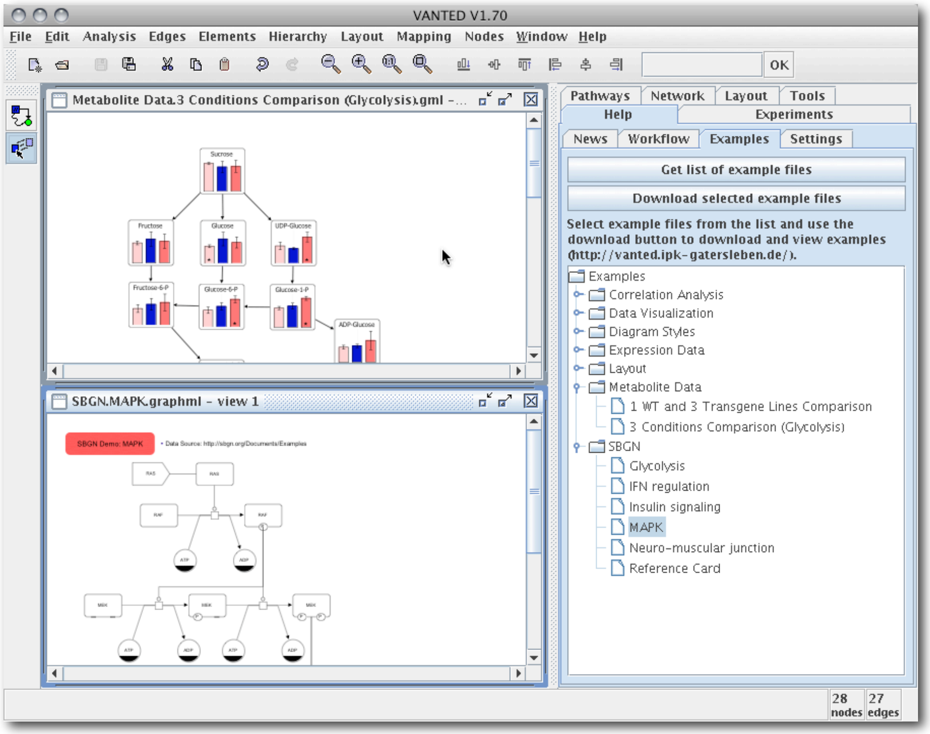
VANTED workshop

IPK 2010

Christian Klukas
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Germany



Select example files from the list and use the download button to download and view examples (http://vanted.ipk-gatersleben.de/).

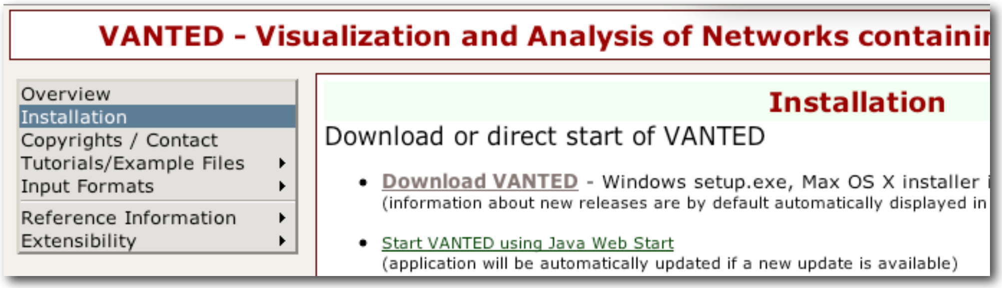
1. Installation and Startup
2. Basic Network Editing
3. Loading of Experiment Data
4. Data Mapping
5. Data Visualization
6. Statistical Analysis
7. Developer Information - Systems Architecture
8. Developer Information - Development of Add-ons

VANTED Tutorial

Visualization and Analysis of Networks containing Experimental Data

HTTP://VANTED.IPK-GATERSLEBEN.DE

1. Installation and Startup

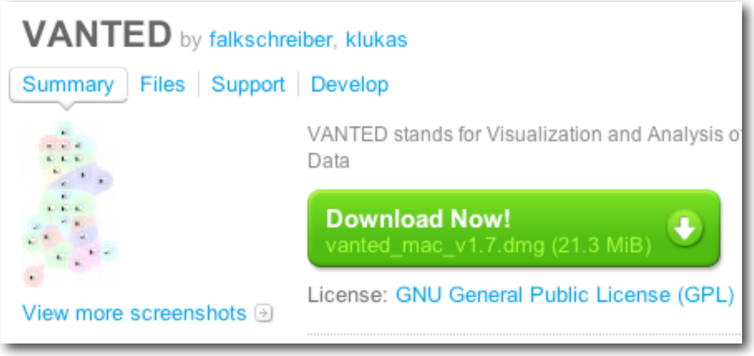


Installation

Download or direct start of VANTED

- Download VANTED - Windows setup.exe, Max OS X installer
  (information about new releases are by default automatically displayed in
- Start VANTED using Java Web Start
  (application will be automatically updated if a new update is available)

Fig. 1: Go to http://vanted.ipk-gatersleben.de, click "Installation" and then "Download VANTED".

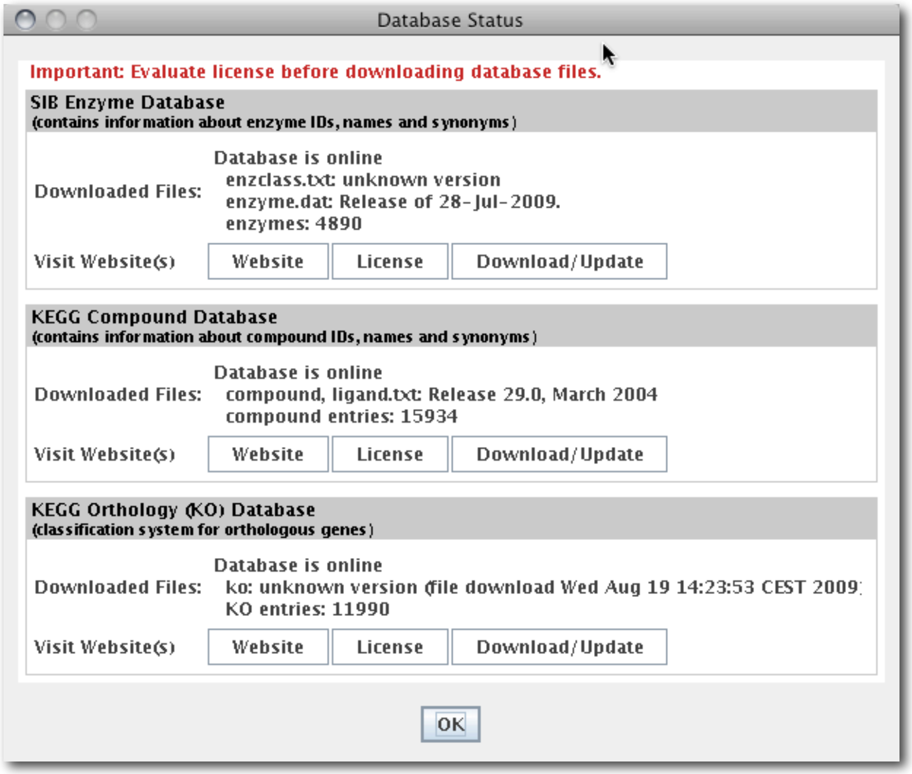


VANTED stands for Visualization and Analysis o
Data

License: GNU General Public License (GPL)

Fig. 2: Click the "Download Now!" button.

Once installed and started, a number of dialogs are shown, which need to be confirmed (to enable KEGG based functionality for academic users) or unconfirmed (commercial users). After the first start, a download dialog will be shown and three download buttons enable access to a number of synonym/annotation-databases. The dialog may be opened at a later time with the menu command "Help/Database Status".



Important: Evaluate license before downloading database files.

SIB Enzyme Database
(contains information about enzyme IDs, names and synonyms)

Downloaded Files: Database is online
enzclass.txt: unknown version
enzyme.dat: Release of 28-Jul-2009.
enzymes: 4890

Visit Website(s)

KEGG Compound Database
(contains information about compound IDs, names and synonyms)

Downloaded Files: Database is online
compound, ligand.txt: Release 29.0, March 2004
compound entries: 15934

Visit Website(s)

KEGG Orthology (KO) Database
(classification system for orthologous genes)

Downloaded Files: Database is online
ko: unknown version (file download Wed Aug 19 14:23:53 CEST 2009
KO entries: 11990

Visit Website(s)

Fig. 3: Click the three "Download/Update" buttons and then "OK".

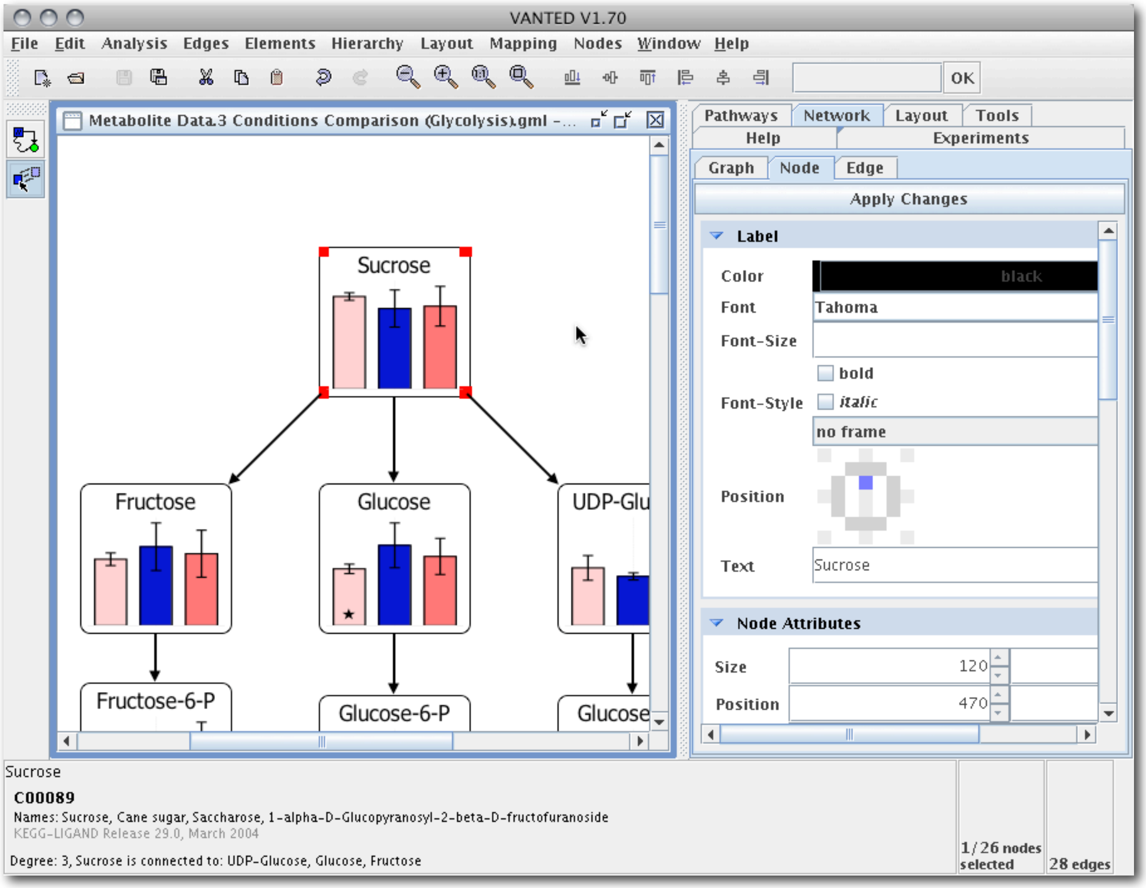
2. Basic Network Editing

Pathways or other kinds of biological networks may be loaded from files and edited, or may be created from scratch using graph editor functions, available at the left toolbar-buttons.

This button activates the Node/Edge-Creation modus. Click onto the window background to create a node, click a node and then another node to create an edge. While creating edges you may click onto the background of the window to create edge bends.

This button activates the Node-Movement modus. It is then possible to select nodes and/or edges (click element or click onto the window background and hold the mouse button while selecting elements). After selecting an edge, click again and hold the mouse to create edge bends. Double click nodes or edges to add labels. Color, shape, label fonts and many more attributes may be modified using the side panel "Network".

Network files may be loaded using drag & drop from the file manager or by using "File/Open". Supported formats are GML, GraphML, Pajek .NET, DOT, KGML, SBML, and more. The default file format is GML, also GraphML retains all characteristics of the graph view. When using other formats, information about visual properties (such as color and size of elements) and other attributes may not be stored.



Color: black
Font: Tahoma
Font-Size:
Font-Style: ☐ bold ☐ *italic* no frame
Position
Text: Sucrose

Size: 120
Position: 470

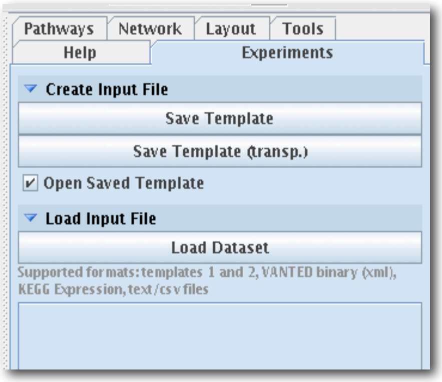
Sucrose
C00089
Names: Sucrose, Cane sugar, Saccharose, 1-alpha-D-Glucopyranosyl-2-beta-D-fructofuranoside
KEGG-LIGAND Release 29.0, March 2004
Degree: 3, Sucrose is connected to: UDP-Glucose, Glucose, Fructose

1/26 nodes selected
28 edges

Fig. 4: Main window, showing Node-Movement and Node/Edge-Creation (active) tool buttons at the left, graph window in the middle, and Node-Edit side panel for editing characteristics of selected elements at the right.

3. Loading of Experiment Data

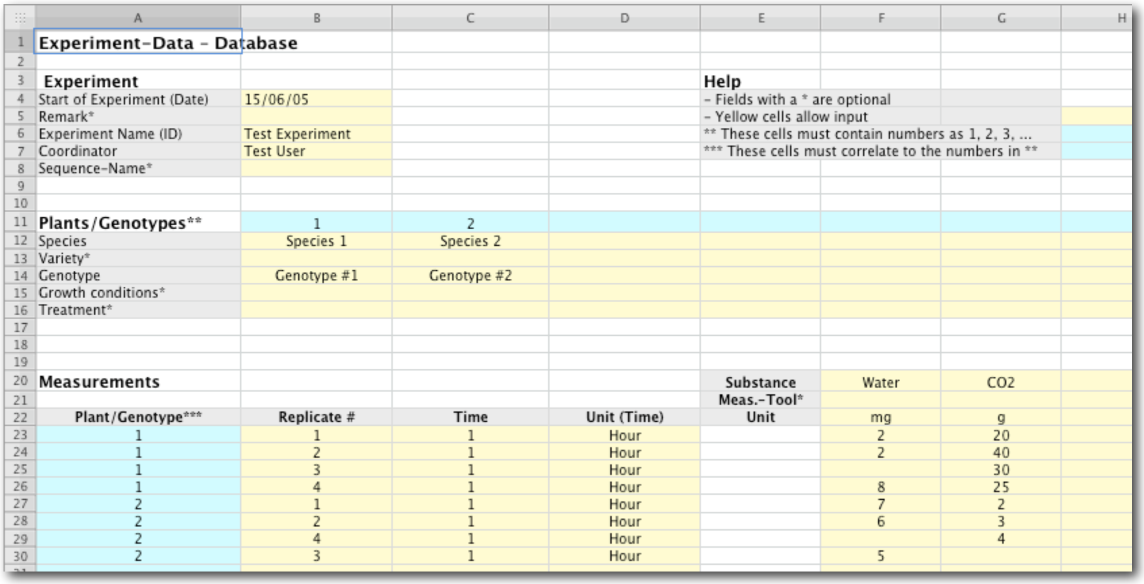
Use the side panel "Experiments" to create a VANTED template file, which may be filled and loaded into the system (Fig. 5). Click "Create Input File" and then use the "Save Template" button to create a VANTED template file, which may be filled with experiment data (Fig. 6).



☑ Open Saved Template

Supported formats: templates 1 and 2, VANTED binary (xml), KEGG Expression, text/csv files

Fig. 5



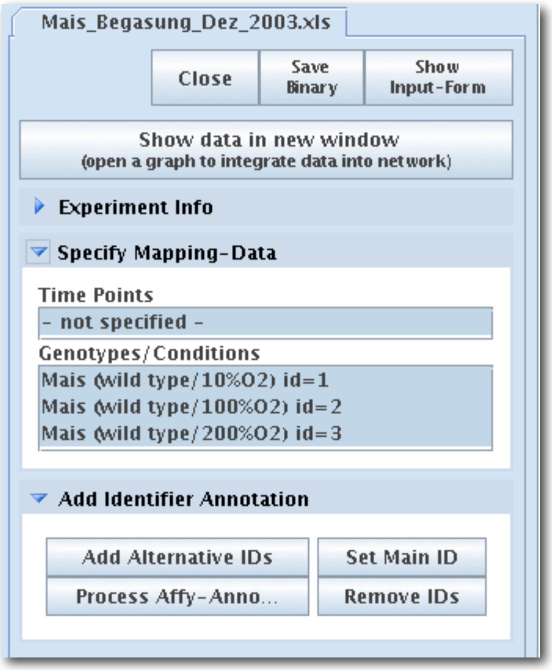
	A	B	C	D	E	F	G	H
1	**Experiment-Data - Database**							
2								
3	**Experiment**				**Help**			
4	Start of Experiment (Date)	15/06/05			- Fields with a * are optional			
5	Remark*				- Yellow cells allow input			
6	Experiment Name (ID)	Test Experiment			** These cells must contain numbers as 1, 2, 3, ...			
7	Coordinator	Test User			*** These cells must correlate to the numbers in **			
8	Sequence-Name*							
9								
10								
11	**Plants/Genotypes****	1	2					
12	Species	Species 1	Species 2					
13	Variety*							
14	Genotype	Genotype #1	Genotype #2					
15	Growth conditions*							
16	Treatment*							
17								
18								
19								
20	**Measurements**				**Substance**	Water	CO2	
21					**Meas.-Tool***			
22	**Plant/Genotype*****	**Replicate #**	**Time**	**Unit (Time)**	**Unit**	mg	g	
23	1	1	1	Hour		2	20	
24	1	2	1	Hour		2	40	
25	1	3	1	Hour			30	
26	1	4	1	Hour		8	25	
27	2	1	1	Hour		7	2	
28	2	2	1	Hour		6	3	
29	2	4	1	Hour			4	
30	2	3	1	Hour		5		

Fig. 6: Experiment data template file, consisting of three sections: top: experiment header, with basic information about the experiment, middle: conditions/genotypes, bottom: measurement data

As an alternative, you may use a "unstructured" data file, where the first column contains the substance identifiers and subsequent columns contain measurement data for different conditions/genotypes, replicates or time points. Upon loading a data file using the "Load Dataset" Button, VANTED will ask for missing information about the experiment and about data columns (only for unstructured data files).

4. Data Mapping

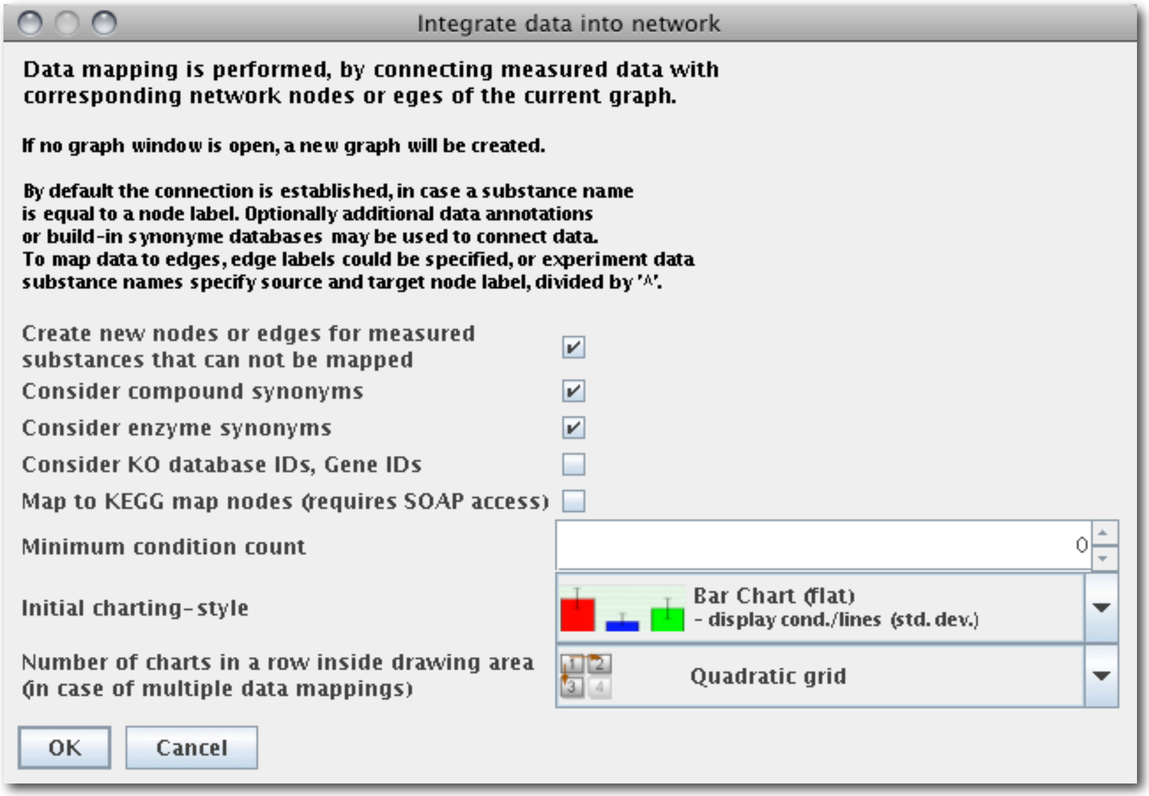
Once a data set is loaded, a tab panel is shown (Fig. 7). In case no network file is open the button on the top "Show data in new window" displays the data of all measured substances. For each substance a graph node is created. Data is shown as a line- or bar-chart, or using color-coding, depending on the number of conditions and time points. If a target network is opened, the top button text reads "Perform data mapping". Once clicked, a dialog window is shown (Fig. 8).



Time Points: - not specified -

Genotypes/Conditions:
Mais (wild type/10%O2) id=1
Mais (wild type/100%O2) id=2
Mais (wild type/200%O2) id=3

Fig. 7



Data mapping is performed, by connecting measured data with corresponding network nodes or eges of the current graph.

If no graph window is open, a new graph will be created.

By default the connection is established, in case a substance name is equal to a node label. Optionally additional data annotations or build-in synonyme databases may be used to connect data.
To map data to edges, edge labels could be specified, or experiment data substance names specify source and target node label, divided by '^'.

Create new nodes or edges for measured substances that can not be mapped ☑
Consider compound synonyms ☑
Consider enzyme synonyms ☑
Consider KO database IDs, Gene IDs ☐
Map to KEGG map nodes (requires SOAP access) ☐
Minimum condition count 0
Initial charting-style: Bar Chart (flat) - display cond./lines (std. dev.)
Number of charts in a row inside drawing area (in case of multiple data mappings): Quadratic grid

Fig. 8: Properties of the "Perform data mapping" button command.

Experiment data is integrated into a network element (node or edge) in case the particular label of the graph element is equal to a substance name of the experiment data set. Additionally, synonyms, extracted from a number of databases, may be considered during the data mapping. For example, by default compound and enzyme synonyms are considered. To map data onto edges, edges need to be labeled, or the measured substance name needs to be specified as "A^B", which means, the data should be mapped onto all edges connecting two nodes labeled as "A" and "B".

5. Data Visualization

During data mapping, an initial charting-style is set. You may choose another kind of data visualization by selecting nodes, which contain data, and by using the side panel "Network" and then "Node". Certain characteristics, which should be the same for all data nodes (e.g. colors for different conditions) can be edited in the "Network/Graph" tab.

Using the menu command "Elements/Evaluate Properties" mapped data and other properties of the graph elements are analyzed and the results are stored as additional attributes of the graph elements. Subsequently, this information may be used, e.g. for search operations ("Edit/Search...") or for modifying graph element colors or sizes (e.g. edge width), by using the commands "Elements/Set Color dep. on Attribute Value" and "Elements/Set Visual Properties dep. on Attribute Value".

6. Statistical Analysis

The side panel "Tools/Statistics" contains a number of functions to compare samples of different conditions and to correlate experiment data of different substances.

7. Developer Information - Systems Architecture

VANTED is based on the extensible graph editor toolkit GRAVISTO, developed at the University of Passau (http://gravisto.fim.uni-passau.de/). VANTED is open source (GPL), implemented in Java and uses a number of external libraries such as JFreeChart for displaying diagrams. More details are available from the "Help/About VANTED" menu.

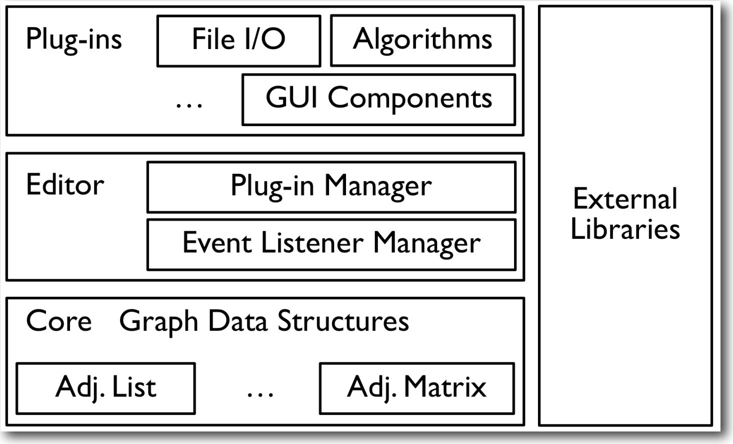


Fig. 9: VANTED system architecture, consisting of three levels, "Core", "Editor" and "Plugins".

Development of the system is done using the Eclipse development environment. The complete source code may be downloaded from SourceForge (http://sourceforge.net/projects/vanted/).

Functions of the "Core" layer are independent of the "Editor" and "Plugins" layers. The "Editor" layer uses functions from the "Core" but not from the "Plugins" layer. The three layers of VANTED (see Fig. 9) correspond to three Eclipse projects, which effectively separates the code bases as needed. The whole system is built around the idea of using Plugins and Managers, therefore aspects such as menu commands, toolbars, side panels, graph attributes, visualization components in the graph view, file filters and much more may be easily extended, exchanged or removed from the system as needed.

8. Developer Information - Development of Add-ons

The code available from SourceForge contains a “Add-on-Example” project which shows how to develop the following types of system extensions:

File name	Kind of extension
ExemplaryAddon.java	this file specifies the features (see below) of the system extension
AddonTab.java	a side panel
DoNothingAlgorithm.java	a menu command which works on the active graph
MyCircleLayout.java	a graph layout algorithm, shown in the side panel “Layout”
MyFirstView.java	a custom graph view
OpenViewAlgorithm.java	a menu command which opens an additional certain kind of view for the current graph
RightClickContextMenuAlgorithm.java	an algorithm which provides right-click context menu items for the graph view

Using the supplied Ant-script “createAdd-on.xml”, a JAR file is created, which may be installed on the users machines (using drag & drop or using the button”Install / Configure Add-ons” in the “Help/Settings” side panel). During development VANTED may be started with the current add-on code base using the main-method contained in the file “StartVantedWithAddon.java”.

[1] Björn H. Junker, Christian Klukas and Falk Schreiber: VANTED: A System for Advanced Data Analysis and Visualization in the Context of Biological Networks. *BMC Bioinformatics* 7: 109.1-13, 2006.