

# NetSeed User Manual

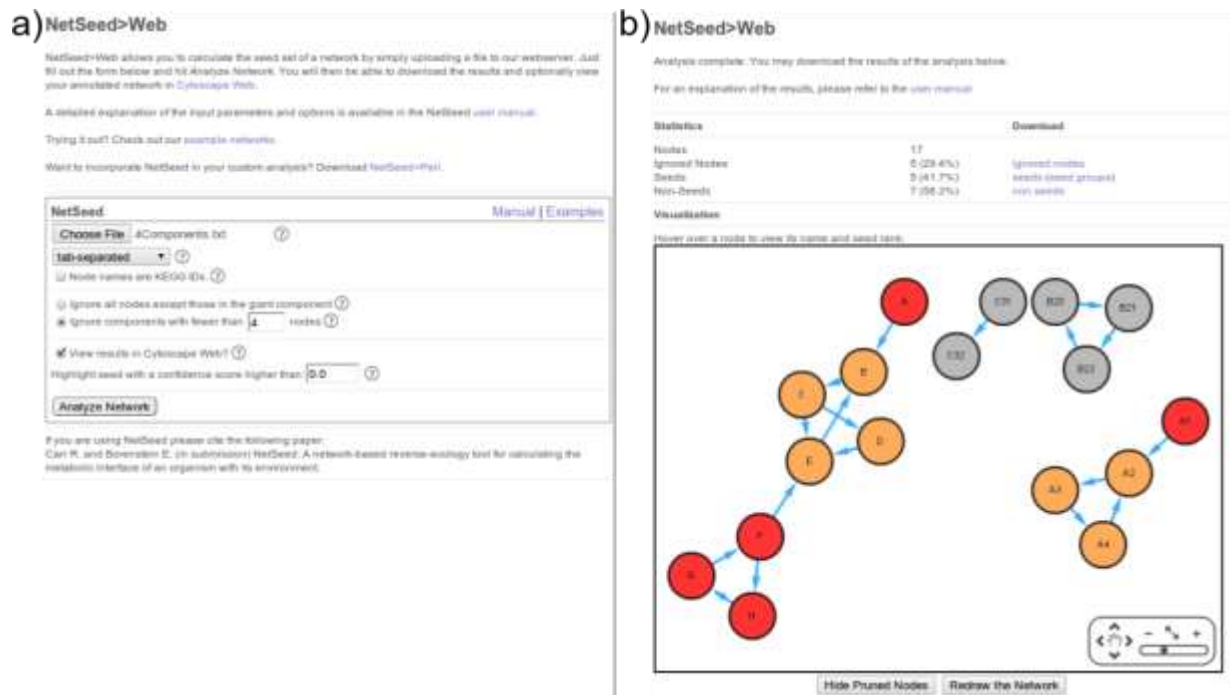
NetSeed is a toolkit for identifying the seed set of networks, available as an online tool and as a Perl module. NetSeed is developed by the Borenstein group at the University of Washington and is available online at <http://elbo.gs.washington.edu/tools/NetSeed/>.

## NetSeed>Web

NetSeed>Web allows researchers to calculate the seed set of a network online and requires only a web browser. The NetSeed>Web application is built on top of the NetSeed>Perl module and has been extended to include visualization and functional analysis options.

### Overview of use

To determine the seed set of a network using NetSeed>Web, the user simply uploads their network file to the NetSeed server by filling out the form shown in Figure 1 a). After checking the integrity of the network file, NetSeed>Web calculates the seed set, the seed groups and non-seed nodes. Links to files containing these node sets, as well as to all the optionally ignored nodes, are displayed with statistics giving the percentage of nodes falling in each category, as shown in Figure 1 b). If the nodes have been annotated with KEGG IDs, a page with links to the KEGG entries is presented, allowing the user to investigate the functional properties of seeds, seed groups, and non-seed nodes. A graphical representation of the network with the seed set highlighted may be shown with Cytoscape Web, allowing the user to examine how the seeds are distributed within the network.



**Figure 1:** The NetSeed>Web interface. a) The user is first presented with a form to upload the network file and select the analysis options. b) The network is analyzed and the results page is displayed, shown here with the optional Cytoscape visualization.

## Setting up the analysis

### Selecting the network file to analyze

**Choose File / Browse:** Select the network file on your local drive. In the drop-down box below, choose the appropriate file type.

NetSeed>Web supports network files formatted as plain text and in SBML, the systems biology markup language. For both formats, node names may be written with any combination of numbers, letters, periods, dashes, underlines, and spaces.

Plain text files have one edge per line with nodes separated by a space, comma or tab. Each edge must use the same separating character. Files that use a space as a separator may not use spaces in the node names. Such files are commonly created with programs like Cytoscape and Excel.

SBML files are used to create networks by reading in the reaction elements. One edge is created for each reactant – product pair.

**Node names are KEGG IDs:** If the node names are taken from the Kyoto Encyclopedia of Genes and Genomes (KEGG), NetSeed>Web can create links to both seeds and non-seeds to their entries in KEGG. If visualization with Cytoscape Web is

enabled, double-clicking on a node in Cytoscape Web will open a new window to the KEGG database entry for the node.

## Analysis Options

Once you have specified the network file, the next step is to choose what parts of the network should be considered in the seed set calculation.

***Ignore all nodes except those in the giant component:*** Selecting this will perform the seed set calculation for only the largest component (connected set of nodes) in your network. All smaller components will be ignored. This is a good option for networks with a dominant component.

***Ignore components with fewer than x nodes:*** Nodes belonging to components with fewer than  $x$  nodes will be ignored. This is a good option for networks that contain many small and trivial components. (Default is 0: Analyze all nodes.)

## Visualization Options

NetSeed>Web offers the option of viewing the results in a Cytoscape Web browser (<http://cytoscapeweb.cytoscape.org>). The Cytoscape Web browser is flash-based program for drawing networks. NetSeed>Web will draw your network in Cytoscape Web with seeds shown in red, non-seed nodes in orange, and ignored nodes in gray. Viewing the results in CytoscapeWeb will not change the analysis. Cytoscape Web is designed for networks with fewer than 2000 elements (nodes and edges), so NetSeed>Web will prompt the user before drawing these networks.

***View results in Cytoscape Web:*** Check to view the results in Cytoscape Web—requires Adobe Flash. (Default is checked.)

***Highlight seeds with a confidence score higher than x:*** Seeds with a confidence score greater than  $x$  will be drawn in red in Cytoscape Web. This option does not do anything if visualization in Cytoscape Web is disabled. (Default is 0: Highlight all seeds.)

## Viewing the Results

NetSeed>Web calculates the set of seeds for the uploaded network and presents the results as a set of files and statistics.

***Nodes:*** The total number of nodes in the analyzed network.

**Ignored Nodes:** The number of nodes (and percent of the total number of nodes) that were not considered. Nodes are ignored if they are in components which are not considered by the analysis (see [Analysis Options](#), above). A file containing the names of ignored nodes is available to download.

**Seeds:** The number of nodes in the seed set (and percent of the total number of nodes). The seed set can be downloaded as a list of all seed nodes and their confidence scores (“seeds”) or as a list of seed groups (“grouped”). A seed group is a set of seeds that are interdependent; in combination they have a confidence score of 1.

**Non-Seeds:** Nodes which were found to not be in the seed set (and percent of the total number of nodes).

**KEGG Annotations:** If *Node names are KEGG IDs* was selected, NetSeed>Web will display link to a file with a list of seeds and non-seeds linking to their entries in KEGG. If visualization with Cytoscape Web is enabled, double-clicking on a node in Cytoscape Web will open a new window to the KEGG database entry for the node.

**Show / Hide Ignored Nodes:** If visualization with Cytoscape Web is enabled and components were omitted from the calculation, this button toggles the visibility of the ignored nodes.

**Redraw Network:** If visualization with Cytoscape Web is enabled, this button forces Cytoscape Web to re-render the network layout.

## NetSeed>Perl

NetSeed>Perl is a stand-alone Perl module that implements the network seed analysis. It is distributed under the GPL and can be readily incorporated into custom analysis tools.

## Installation Instructions

To install NetSeed Perl, simply download the package from <http://depts.washington.edu/elbogs/NetSeed/NetSeedPerl.zip>. This is zip archive contains the following files:

*NetSeed.pm*: The NetSeed Perl module.

*example.pl*: A Perl script demonstrating how to use the NetSeed Perl module.

*COPYING*: A copy of the GNU General Public License. This is required to be distributed with the NetSeed>Perl package.

NetSeed.pm can be placed in any directory that can be accessed from a Perl script (for example, with ``use lib'`).

## Interface

The NetSeed>Perl module handles all calculations internally. NetSeed>Perl offers an interface to the network seed set calculation through a set of functions imported with the package.

**CalculateSeeds** (fileHandle, separator, giantComponentFlag, minimalComponentSize)

### Inputs

*fileHandle*: An open fileHandle to a file containing the network.

The network file should be a plain text file with an edge on each line. An edge is written as two node names separated by a single character. Each edge must use the same separating character, which may not be used in the node names.

*separator*: The character used to separate the columns in the network file.

*giantComponentFlag*: 0 or 1. 1 specifies that only the giant component of the network should be considered.

*minimalComponentSize*: digit. The minimal size of component to be considered. Components with fewer nodes than \$minimalComponentSize will be ignored.

Outputs: none.

### **Seeds ()**

Inputs: None.

Outputs: A hash containing the calculated network seeds. The keys of the hash are the seed node names and values are the seed confidence scores.

### **GroupedSeeds ()**

Inputs: None.

Outputs: A hash containing the network seeds grouped by interdependence. The keys of the hash contain each seed group, with seed names separated by \$separator.

### **NonSeeds ()**

Inputs: None.

Outputs: A hash containing the nodes in the network that are not seeds. The keys of the hash contain the node names.

### **IgnoredNodes ()**

Inputs: None.

Outputs: A hash containing the nodes in the network that were ignored by the NetSeed calculation. The keys of the hash contain the node names.

### **AllNodes ()**

Inputs: None.

Outputs: A hash containing all the nodes in the network. The keys of the hash contain the node names.

### **AllEdges ()**

Inputs: None.

Outputs: A two-dimensional hash containing all the edges in the network. The keys of the hash contain the source node names. The value of each key is a hash containing the names of its target nodes.

### **NumElements ()**

Inputs: None.

Outputs: An integer containing the number of elements (nodes + edges) in the network.

## **Example Use**

See ***example.pl*** in the NetSeed distribution for an example of how to use NetSeed>Perl in a custom Perl script.