

4.1 Sci2 Release Notes v0.5 alpha

Announcement

The Cyberinfrastructure for Network Science Center is pleased to announce the release of the Sci2 (Science of Science) Tool v0.5 alpha, which contains much of the functionality of the Network Workbench tool as well as a variety of new general purpose and scientometrics-focused algorithms. Sci2 is still undergoing significant changes, but already has many new features that are worth exploring. Please visit the [new Sci2 homepage](#) to download and learn more about Sci2 .

Sci2 is well documented with over [12 tutorials](#) prepared for NIH in Summer 2010 and an [online user manual](#) with more than 100 pages of content, including [sample Sci2 workflows](#) and [documentation for every algorithm in Sci2](#). New features in Sci2 so far include web-based Yahoo! and desktop Geocoders, two different ways to overlaying geographic information on U.S. and World Maps, customizable stop word lists, and an algorithm for combining two networks into one. We are also developing a new scalable database-oriented scientometrics pipeline capable of producing many new types of networks and tables based on ISI, NSF, and now generic CSV data. This new pipeline can be downloaded as a supplemental preview package from the download section of the [Sci2 site](#), and will be available by default in Sci2 for future releases. We have also fixed many bugs from Network Workbench and previous Sci2 versions. See below for a full list of changes since Sci2 v0.3.

Network Workbench and Sci2 are both based on the [CIShell framework](#), which makes it easy to add new algorithms to each tool. See our recent video featured in the [Communications of the ACM](#) or the [CIShell wiki](#) for more information.

Documentation

[Download an offline copy of this manual \(87.6 MB\) \(Updated May 31, 2011\)](#)

[Download an offline copy of the CIShell Manual, which contains algorithm documentation for Sci2 \(19.9 MB\) \(Updated May 31, 2011\)](#)

Tool Change Log

New Features

- Yahoo! Geocoder.
- Combine 2 Networks algorithm.
- Added support for custom stop word lists (the stop word algorithm now refers to an external file, which can be copied and edited to contain different stop words).
- Added GUESS support for saving out node positions. This can be used in combination with network timeslicing to help in the creation of an animation or a series of network images where the network grows over time.
- GUESS zoom level can now be adjusted in absolute terms. This helps when trying to compare to GUESS visualizations (set the zoom levels to agree).
- Added support for loading files by dragging them from the desktop to the data manager, and saving files via dragging them out of the Data Manager.
- Added support for loading multiple files simultaneously via File -> Load.
- [New website](#), including the 'Ask an Expert' feature.
- In supplemental database package:
 - New 'Extract X by Year' algorithms that work as input for Burst detection.
 - database dump algorithm
 - bibliographic coupling network extractions
 - longitudinal summary extraction
 - Added support for generic-csv databases, including the loader GUI and the extraction GUI.

Improvements

- Improved support for UTF-8 data.
- Shortened file names for loaded files.
- improved zip code parsing.
- Fixed circle sizes in GeoMaps so they could be adjusted by a scaling factor.
- Improved GeoMaps so it only shows projections that make sense in the context of the map you want to show (either World or U.S. map).
- Enhanced Geocoder so it can resolve congressional districts to geographic coordinates.
- Improved CSV reader so it could handle some new variations in CSV input (including data from the NIH RePORTER site).
- GUESS now shows in its title bar the name of the network it is displaying.
- Blondel Community Detection now handles non-integer weights.
- In supplemental database package:
 - Expanded database default list of journal names to merge.
 - Improved database performance
 - Improved performance with bibliographic coupling extraction.
 - Database loaders now show % complete when loading data.
 - Improved database loading performance.
 - Improved database co-authorship extraction to include earliest & most recent publication date for each author node.
 - Added 'total # works' and 'total times cited' counts to the author extraction table.
 - Changed python scripts and/or database and old-pipeline attribute names so our custom GUESS python scripts work for both database and old-pipeline output.

Bug fixes

- Fixed a text display error on Macs.
- Fixed an issue where GUESS couldn't display certain pajek.net files.

- Fixed an issue where filenames could start or end with a space.
- Fixed an error with creating directed networks from some NSF files.
- Fixed an issue where text-based pipeline co-authorship network extraction no longer merged names with different cases.
- Fixed an issue in GUESS where certain drop-down boxes would not be populated.
- Fixed an issue with dates in the NSF database.
- Fixed an issue where Extract Authors by Year output would sometimes be invalid.
- Fixed issue with having multiple files with the same name in the Data Manager.
- Fixed an issue with the ISI paper -> reference matcher.
- Fixed a bug where Node Betweenness Centrality would return a graph with zero nodes.
- Fixed an error condition in Co-Word Occurrence algorithm.
- Fixed a confusing package naming issue between two versions of delete isolates.
- Fixed merging capabilities in non-db pipeline.
- Fixed broken links under Help -> About.
- Fixed error handling in Geo Maps when an input file had no numeric columns (and hence no basis for size or color coding).
- Fixed an issue with 'delete isolates' and 'weak component clustering'.
- Fixed an issue with NWB formatting that was causing issues in GUESS.
- Fixed an error with handling EndNote data.
- Fixed a number of errors in GUESS python scripts.
- Fixed an error in 'Extract Nodes Above or Below Value'.
- Fixed an error in 'normalize text'.
- Fixed an issue in GUESS where edges in UTF-8 files would not be shown.
- Fixed old references in Sci2 that pointed to the old bug tracker to point to our new bug tracker.
- Fixed an issue with the output of document citation network where it wouldn't work with Weighted Page Rank.
- Fixed issue where some pajek.net files would not load properly in pajek.
- Changed some text that still said 'NWB' to 'Sci2'.
- Fixed an issue with viewing pajek.net files immediately after they were loaded into Sci2.
- Fixed an issue with GUESS colorizing.
- Fixed a merging error produced by one of our Sci2 workflows.
- Fixed an issue with Node Betweenness Centrality.
- Fixed an issue with normalizing author names.
- Fixed an issue where Aggregate Data would not have a good error message when columns have null values in them.
- Fixed an issue with viewing the CSV output of Extract ZIP code.
- Fixed an issue with GUESS displaying XGMML input strangely.
- Fixed an issue with the Network Analysis Toolkit where it was sensitive to whitespace in pajek.net files.
- Fixed an issue with Delete Isolates where the number of nodes deleted differed from what it was reporting on the Console.
- In supplemental database package:
 - Fixed a database error with Core journal co-citation output producing strange results in NAT and GUESS.
 - Fixed a database error where Extract Authors would sometimes return zero results.
 - Fixed an issue with loading certain unusual ISI files into the ISI database.
 - Changed text so that it always used the more inclusive term 'Document Sources' rather than 'Journals' where appropriate.
 - Fixed a bug where database produced bad output for GUESS.