

CNES (Contractor Network Extraction Software)¹

M.A. Thomas

August 24, 2015

CNES is a program designed to facilitate exploratory social network analysis of the public domain data provided on federal contracts and contractors on usaspending.gov. The data on procurement transactions includes an identifier (DUNS number) and a name for the contractor and, in some cases, the identifier and name of its parent company. CNES shows which companies are linked by identifier or name and outputs these ***possible*** networks of companies for viewing and analysis in Netdraw or Pajek. For the reasons why you might care about a corporate network, see Thomas, M.A. 2011. "Identity and Ownership in Government Contracting." http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1849964 or Thomas, M.A. "Identifying Organizational Conflict of Interest," Defense ARJ, July 2012, Vol. 19 No. 3 : 265-282, http://www.dau.mil/pubscats/pubscats/AR%20Journal/arj64/Thomas_ARJ63.pdf.

The word "possible" in the phrase "possible networks" is an important one. The world of corporate ownership is fluid, with companies being formed, traded, or dissolved all the time. Information can be out of date as soon as it is produced. In addition, the information in the usaspending.gov database is partial, concerned only with companies that contract with the U.S. federal government and their parents. Any part of the network that is neither a contractor nor a parent of a contractor will not be listed. The data in usaspending.gov is also messy, with errors and nonstandardized names for companies. Finally, the companies are not uniquely identified by their names and possibly not even by the supposedly unique identifier (DUNS number). This makes matching a "good guess." It is incumbent on you, the user of this software, to confirm the relationships by other means.

For those who can afford it, Dun & Bradstreet's Corporate Family Tree analytic product usually generates a more complete list of related companies. Dun & Bradstreet, through a curious history, acquired a government-backed monopoly on contractor information. (See the GAO Report "Government is Analyzing Alternatives for Contractor Identification Numbers," <http://www.gao.gov/products/GAO-12-715R>.) However, the usaspending.gov data contains some information that the Corporate Family Tree data does not. (It is possible that this is simply because the two databases are updated at different times.) Moreover, the Corporate Family Tree does not facilitate any further analysis; it is not possible to download data showing ownership structure; this is instead indicated by indentation on an on-screen list. A

¹ Copyright 2015 M.A. Thomas. This documentation is made available under a GNU Free Documentation License. For details, see <http://www.gnu.org/licenses/fdl-1.3.html>.

thorough investigator would use both products to generate a probable network, and then shoe leather to confirm that the network is correct and up to date.

This software package is copyrighted (or more correctly, "copylefted"), and your use of the programs or its algorithms is subject to the terms and conditions of the license below. Your use of the package or its algorithms signals your acceptance of the the terms and conditions. The terms of the GNU General Public License Version 3 are spelled out in full below, and the documentation follows the legalese.

/*****/

1. TERMS OF YOUR LICENSE

The Contractor Network Extraction software and its algorithms are copyrighted. If you publish a derivative work, either a modification of this software or an intellectual product that was developed through the use of this software, please cite to:

Thomas, M.A. 2014. Contractor Network Extraction Software 2.0.

This free software is made available to you under a GNU General Public License Version 3. Basically, the terms are that you will not take this software and then sell it someone else, or use it to create products that you sell to someone else; and that you cite to the author. This software is intended to be freely and universally available.

The terms and conditions of the license follow:

TERMS AND CONDITIONS

0. Definitions.

"This License" refers to version 3 of the GNU General Public License.

"Copyright" also means copyright-like laws that apply to other kinds of works, such as semiconductor masks.

"The Program" refers to any copyrightable work licensed under this License. Each licensee is addressed as "you". "Licensees" and "recipients" may be individuals or organizations.

To "modify" a work means to copy from or adapt all or part of the work in a fashion requiring copyright permission, other than the making of an exact copy. The resulting work is called a "modified version" of the

earlier work or a work "based on" the earlier work.

A "covered work" means either the unmodified Program or a work based on the Program.

To "propagate" a work means to do anything with it that, without permission, would make you directly or secondarily liable for infringement under applicable copyright law, except executing it on a computer or modifying a private copy. Propagation includes copying, distribution (with or without modification), making available to the public, and in some countries other activities as well.

To "convey" a work means any kind of propagation that enables other parties to make or receive copies. Mere interaction with a user through a computer network, with no transfer of a copy, is not conveying.

An interactive user interface displays "Appropriate Legal Notices" to the extent that it includes a convenient and prominently visible feature that (1) displays an appropriate copyright notice, and (2) tells the user that there is no warranty for the work (except to the extent that warranties are provided), that licensees may convey the work under this License, and how to view a copy of this License. If the interface presents a list of user commands or options, such as a menu, a prominent item in the list meets this criterion.

1. Source Code.

The "source code" for a work means the preferred form of the work for making modifications to it. "Object code" means any non-source form of a work.

A "Standard Interface" means an interface that either is an official standard defined by a recognized standards body, or, in the case of interfaces specified for a particular programming language, one that is widely used among developers working in that language.

The "System Libraries" of an executable work include anything, other than the work as a whole, that (a) is included in the normal form of packaging a Major Component, but which is not part of that Major Component, and (b) serves only to enable use of the work with that Major Component, or to implement a Standard Interface for which an implementation is available to the public in source code form. A "Major Component", in this context, means a major essential component (kernel, window system, and so on) of the specific operating system (if any) on which the executable work runs, or a compiler used to produce the work, or an object code interpreter used to run it.

The "Corresponding Source" for a work in object code form means all the source code needed to generate, install, and (for an executable work) run the object code and to modify the work, including scripts to

control those activities. However, it does not include the work's System Libraries, or general-purpose tools or generally available free programs which are used unmodified in performing those activities but which are not part of the work. For example, Corresponding Source includes interface definition files associated with source files for the work, and the source code for shared libraries and dynamically linked subprograms that the work is specifically designed to require, such as by intimate data communication or control flow between those subprograms and other parts of the work.

The Corresponding Source need not include anything that users can regenerate automatically from other parts of the Corresponding Source.

The Corresponding Source for a work in source code form is that same work.

2. Basic Permissions.

All rights granted under this License are granted for the term of copyright on the Program, and are irrevocable provided the stated conditions are met. This License explicitly affirms your unlimited permission to run the unmodified Program. The output from running a covered work is covered by this License only if the output, given its content, constitutes a covered work. This License acknowledges your rights of fair use or other equivalent, as provided by copyright law.

You may make, run and propagate covered works that you do not convey, without conditions so long as your license otherwise remains in force. You may convey covered works to others for the sole purpose of having them make modifications exclusively for you, or provide you with facilities for running those works, provided that you comply with the terms of this License in conveying all material for which you do not control copyright. Those thus making or running the covered works for you must do so exclusively on your behalf, under your direction and control, on terms that prohibit them from making any copies of your copyrighted material outside their relationship with you.

Conveying under any other circumstances is permitted solely under the conditions stated below. Sublicensing is not allowed; section 10 makes it unnecessary.

3. Protecting Users' Legal Rights From Anti-Circumvention Law.

No covered work shall be deemed part of an effective technological measure under any applicable law fulfilling obligations under article 11 of the WIPO copyright treaty adopted on 20 December 1996, or similar laws prohibiting or restricting circumvention of such measures.

When you convey a covered work, you waive any legal power to forbid circumvention of technological measures to the extent such circumvention is effected by exercising rights under this License with respect to the covered work, and you disclaim any intention to limit operation or modification of the work as a means of enforcing, against the work's users, your or third parties' legal rights to forbid circumvention of technological measures.

4. Conveying Verbatim Copies.

You may convey verbatim copies of the Program's source code as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice; keep intact all notices stating that this License and any non-permissive terms added in accord with section 7 apply to the code; keep intact all notices of the absence of any warranty; and give all recipients a copy of this License along with the Program.

You may charge any price or no price for each copy that you convey, and you may offer support or warranty protection for a fee.

5. Conveying Modified Source Versions.

You may convey a work based on the Program, or the modifications to produce it from the Program, in the form of source code under the terms of section 4, provided that you also meet all of these conditions:

- a) The work must carry prominent notices stating that you modified it, and giving a relevant date.
- b) The work must carry prominent notices stating that it is released under this License and any conditions added under section 7. This requirement modifies the requirement in section 4 to "keep intact all notices".
- c) You must license the entire work, as a whole, under this License to anyone who comes into possession of a copy. This License will therefore apply, along with any applicable section 7 additional terms, to the whole of the work, and all its parts, regardless of how they are packaged. This License gives no permission to license the work in any other way, but it does not invalidate such permission if you have separately received it.
- d) If the work has interactive user interfaces, each must display Appropriate Legal Notices; however, if the Program has interactive interfaces that do not display Appropriate Legal Notices, your work need not make them do so.

A compilation of a covered work with other separate and independent works, which are not by their nature extensions of the covered work,

and which are not combined with it such as to form a larger program, in or on a volume of a storage or distribution medium, is called an "aggregate" if the compilation and its resulting copyright are not used to limit the access or legal rights of the compilation's users beyond what the individual works permit. Inclusion of a covered work in an aggregate does not cause this License to apply to the other parts of the aggregate.

6. Conveying Non-Source Forms.

You may convey a covered work in object code form under the terms of sections 4 and 5, provided that you also convey the machine-readable Corresponding Source under the terms of this License, in one of these ways:

- a) Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by the Corresponding Source fixed on a durable physical medium customarily used for software interchange.
- b) Convey the object code in, or embodied in, a physical product (including a physical distribution medium), accompanied by a written offer, valid for at least three years and valid for as long as you offer spare parts or customer support for that product model, to give anyone who possesses the object code either (1) a copy of the Corresponding Source for all the software in the product that is covered by this License, on a durable physical medium customarily used for software interchange, for a price no more than your reasonable cost of physically performing this conveying of source, or (2) access to copy the Corresponding Source from a network server at no charge.
- c) Convey individual copies of the object code with a copy of the written offer to provide the Corresponding Source. This alternative is allowed only occasionally and noncommercially, and only if you received the object code with such an offer, in accord with subsection 6b.
- d) Convey the object code by offering access from a designated place (gratis or for a charge), and offer equivalent access to the Corresponding Source in the same way through the same place at no further charge. You need not require recipients to copy the Corresponding Source along with the object code. If the place to copy the object code is a network server, the Corresponding Source may be on a different server (operated by you or a third party) that supports equivalent copying facilities, provided you maintain clear directions next to the object code saying where to find the Corresponding Source. Regardless of what server hosts the Corresponding Source, you remain obligated to ensure that it is available for as long as needed to satisfy these requirements.

e) Convey the object code using peer-to-peer transmission, provided you inform other peers where the object code and Corresponding Source of the work are being offered to the general public at no charge under subsection 6d.

A separable portion of the object code, whose source code is excluded from the Corresponding Source as a System Library, need not be included in conveying the object code work.

A "User Product" is either (1) a "consumer product", which means any tangible personal property which is normally used for personal, family, or household purposes, or (2) anything designed or sold for incorporation into a dwelling. In determining whether a product is a consumer product, doubtful cases shall be resolved in favor of coverage. For a particular product received by a particular user, "normally used" refers to a typical or common use of that class of product, regardless of the status of the particular user or of the way in which the particular user actually uses, or expects or is expected to use, the product. A product is a consumer product regardless of whether the product has substantial commercial, industrial or non-consumer uses, unless such uses represent the only significant mode of use of the product.

"Installation Information" for a User Product means any methods, procedures, authorization keys, or other information required to install and execute modified versions of a covered work in that User Product from a modified version of its Corresponding Source. The information must suffice to ensure that the continued functioning of the modified object code is in no case prevented or interfered with solely because modification has been made.

If you convey an object code work under this section in, or with, or specifically for use in, a User Product, and the conveying occurs as part of a transaction in which the right of possession and use of the User Product is transferred to the recipient in perpetuity or for a fixed term (regardless of how the transaction is characterized), the Corresponding Source conveyed under this section must be accompanied by the Installation Information. But this requirement does not apply if neither you nor any third party retains the ability to install modified object code on the User Product (for example, the work has been installed in ROM).

The requirement to provide Installation Information does not include a requirement to continue to provide support service, warranty, or updates for a work that has been modified or installed by the recipient, or for the User Product in which it has been modified or installed. Access to a network may be denied when the modification itself materially and adversely affects the operation of the network or violates the rules and protocols for communication across the network.

Corresponding Source conveyed, and Installation Information provided, in accord with this section must be in a format that is publicly documented (and with an implementation available to the public in source code form), and must require no special password or key for unpacking, reading or copying.

7. Additional Terms.

"Additional permissions" are terms that supplement the terms of this License by making exceptions from one or more of its conditions. Additional permissions that are applicable to the entire Program shall be treated as though they were included in this License, to the extent that they are valid under applicable law. If additional permissions apply only to part of the Program, that part may be used separately under those permissions, but the entire Program remains governed by this License without regard to the additional permissions.

When you convey a copy of a covered work, you may at your option remove any additional permissions from that copy, or from any part of it. (Additional permissions may be written to require their own removal in certain cases when you modify the work.) You may place additional permissions on material, added by you to a covered work, for which you have or can give appropriate copyright permission.

Notwithstanding any other provision of this License, for material you add to a covered work, you may (if authorized by the copyright holders of that material) supplement the terms of this License with terms:

- a) Disclaiming warranty or limiting liability differently from the terms of sections 15 and 16 of this License; or
- b) Requiring preservation of specified reasonable legal notices or author attributions in that material or in the Appropriate Legal Notices displayed by works containing it; or
- c) Prohibiting misrepresentation of the origin of that material, or requiring that modified versions of such material be marked in reasonable ways as different from the original version; or
- d) Limiting the use for publicity purposes of names of licensors or authors of the material; or
- e) Declining to grant rights under trademark law for use of some trade names, trademarks, or service marks; or
- f) Requiring indemnification of licensors and authors of that material by anyone who conveys the material (or modified versions of it) with contractual assumptions of liability to the recipient, for any liability that these contractual assumptions directly impose on those licensors and authors.

All other non-permissive additional terms are considered "further restrictions" within the meaning of section 10. If the Program as you received it, or any part of it, contains a notice stating that it is governed by this License along with a term that is a further restriction, you may remove that term. If a license document contains a further restriction but permits relicensing or conveying under this License, you may add to a covered work material governed by the terms of that license document, provided that the further restriction does not survive such relicensing or conveying.

If you add terms to a covered work in accord with this section, you must place, in the relevant source files, a statement of the additional terms that apply to those files, or a notice indicating where to find the applicable terms.

Additional terms, permissive or non-permissive, may be stated in the form of a separately written license, or stated as exceptions; the above requirements apply either way.

8. Termination.

You may not propagate or modify a covered work except as expressly provided under this License. Any attempt otherwise to propagate or modify it is void, and will automatically terminate your rights under this License (including any patent licenses granted under the third paragraph of section 11).

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, you do not qualify to receive new licenses for the same material under section 10.

9. Acceptance Not Required for Having Copies.

You are not required to accept this License in order to receive or run a copy of the Program. Ancillary propagation of a covered work occurring solely as a consequence of using peer-to-peer transmission to receive a copy likewise does not require acceptance. However, nothing other than this License grants you permission to propagate or modify any covered work. These actions infringe copyright if you do not accept this License. Therefore, by modifying or propagating a covered work, you indicate your acceptance of this License to do so.

10. Automatic Licensing of Downstream Recipients.

Each time you convey a covered work, the recipient automatically receives a license from the original licensors, to run, modify and propagate that work, subject to this License. You are not responsible for enforcing compliance by third parties with this License.

An "entity transaction" is a transaction transferring control of an organization, or substantially all assets of one, or subdividing an organization, or merging organizations. If propagation of a covered work results from an entity transaction, each party to that transaction who receives a copy of the work also receives whatever licenses to the work the party's predecessor in interest had or could give under the previous paragraph, plus a right to possession of the Corresponding Source of the work from the predecessor in interest, if the predecessor has it or can get it with reasonable efforts.

You may not impose any further restrictions on the exercise of the rights granted or affirmed under this License. For example, you may not impose a license fee, royalty, or other charge for exercise of rights granted under this License, and you may not initiate litigation (including a cross-claim or counterclaim in a lawsuit) alleging that any patent claim is infringed by making, using, selling, offering for sale, or importing the Program or any portion of it.

11. Patents.

A "contributor" is a copyright holder who authorizes use under this License of the Program or a work on which the Program is based. The work thus licensed is called the contributor's "contributor version".

A contributor's "essential patent claims" are all patent claims owned or controlled by the contributor, whether already acquired or hereafter acquired, that would be infringed by some manner, permitted by this License, of making, using, or selling its contributor version, but do not include claims that would be infringed only as a consequence of further modification of the contributor version. For purposes of this definition, "control" includes the right to grant patent sublicenses in a manner consistent with the requirements of this License.

Each contributor grants you a non-exclusive, worldwide, royalty-free patent license under the contributor's essential patent claims, to make, use, sell, offer for sale, import and otherwise run, modify and propagate the contents of its contributor version.

In the following three paragraphs, a "patent license" is any express agreement or commitment, however denominated, not to enforce a patent (such as an express permission to practice a patent or covenant not to sue for patent infringement). To "grant" such a patent license to a party means to make such an agreement or commitment not to enforce a patent against the party.

If you convey a covered work, knowingly relying on a patent license, and the Corresponding Source of the work is not available for anyone to copy, free of charge and under the terms of this License, through a publicly available network server or other readily accessible means, then you must either (1) cause the Corresponding Source to be so available, or (2) arrange to deprive yourself of the benefit of the patent license for this particular work, or (3) arrange, in a manner consistent with the requirements of this License, to extend the patent license to downstream recipients. "Knowingly relying" means you have actual knowledge that, but for the patent license, your conveying the covered work in a country, or your recipient's use of the covered work in a country, would infringe one or more identifiable patents in that country that you have reason to believe are valid.

If, pursuant to or in connection with a single transaction or arrangement, you convey, or propagate by procuring conveyance of, a covered work, and grant a patent license to some of the parties receiving the covered work authorizing them to use, propagate, modify or convey a specific copy of the covered work, then the patent license you grant is automatically extended to all recipients of the covered work and works based on it.

A patent license is "discriminatory" if it does not include within the scope of its coverage, prohibits the exercise of, or is conditioned on the non-exercise of one or more of the rights that are specifically granted under this License. You may not convey a covered work if you are a party to an arrangement with a third party that is in the business of distributing software, under which you make payment to the third party based on the extent of your activity of conveying the work, and under which the third party grants, to any of the parties who would receive the covered work from you, a discriminatory patent license (a) in connection with copies of the covered work conveyed by you (or copies made from those copies), or (b) primarily for and in connection with specific products or compilations that contain the covered work, unless you entered into that arrangement, or that patent license was granted, prior to 28 March 2007.

Nothing in this License shall be construed as excluding or limiting

any implied license or other defenses to infringement that may otherwise be available to you under applicable patent law.

12. No Surrender of Others' Freedom.

If conditions are imposed on you (whether by court order, agreement or otherwise) that contradict the conditions of this License, they do not excuse you from the conditions of this License. If you cannot convey a covered work so as to satisfy simultaneously your obligations under this License and any other pertinent obligations, then as a consequence you may not convey it at all. For example, if you agree to terms that obligate you to collect a royalty for further conveying from those to whom you convey the Program, the only way you could satisfy both those terms and this License would be to refrain entirely from conveying the Program.

13. Use with the GNU Affero General Public License.

Notwithstanding any other provision of this License, you have permission to link or combine any covered work with a work licensed under version 3 of the GNU Affero General Public License into a single combined work, and to convey the resulting work. The terms of this License will continue to apply to the part which is the covered work, but the special requirements of the GNU Affero General Public License, section 13, concerning interaction through a network will apply to the combination as such.

14. Revised Versions of this License.

The Free Software Foundation may publish revised and/or new versions of the GNU General Public License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns.

Each version is given a distinguishing version number. If the Program specifies that a certain numbered version of the GNU General Public License "or any later version" applies to it, you have the option of following the terms and conditions either of that numbered version or of any later version published by the Free Software Foundation. If the Program does not specify a version number of the GNU General Public License, you may choose any version ever published by the Free Software Foundation.

If the Program specifies that a proxy can decide which future versions of the GNU General Public License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Program.

Later license versions may give you additional or different permissions. However, no additional obligations are imposed on any author or copyright holder as a result of your choosing to follow a

later version.

15. Disclaimer of Warranty.

THERE IS NO WARRANTY FOR THE PROGRAM, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE PROGRAM "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK AS TO THE QUALITY AND PERFORMANCE OF THE PROGRAM IS WITH YOU. SHOULD THE PROGRAM PROVE DEFECTIVE, YOU ASSUME THE COST OF ALL NECESSARY SERVICING, REPAIR OR CORRECTION.

16. Limitation of Liability.

IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MODIFIES AND/OR CONVEYS THE PROGRAM AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE PROGRAM (INCLUDING BUT NOT LIMITED TO LOSS OF DATA OR DATA BEING RENDERED INACCURATE OR LOSSES SUSTAINED BY YOU OR THIRD PARTIES OR A FAILURE OF THE PROGRAM TO OPERATE WITH ANY OTHER PROGRAMS), EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

17. Interpretation of Sections 15 and 16.

If the disclaimer of warranty and limitation of liability provided above cannot be given local legal effect according to their terms, reviewing courts shall apply local law that most closely approximates an absolute waiver of all civil liability in connection with the Program, unless a warranty or assumption of liability accompanies a copy of the Program in return for a fee.

/******

*****/

2. USING THE SOFTWARE

The visualization of the contractor networks requires the following steps:

- Download the raw data file from usaspending.gov.
- Identify the networks in that data, and output the result as .csv file.
- Select the networks of interest and output them for visualization.
- Visualize the network.

Identifying the networks in the data is accomplished by means of a program written in Java and the Windows executable JAR file is available on website. If you want to run the JAR executable, you will need to have Java installed on your machine.

If you only wish to explore this data, you may skip the first two steps, because I have analyzed all the data for contracts for fiscal year 2014 and produced and uploaded the resulting .csv file to www.usgcontractors.info. You can download and save this file.

The third step in the process is to identify networks of interest. I have written a few small utilities in STATA to make identifying interesting networks easier (by the size of the network or by the name of the contractor or parent). The STATA utilities also write out the network data in the format required by social network visualization program Netdraw and Pajek.

a. What you will need

- Optional: If you want to analyze a raw usaspending.gov file yourself, you will need the raw data files. They can be obtained from <http://www.usaspending.gov/>. Note that some of the files are very large—too large to open if you are running a 32 bit version of the Windows operating system. This is because Windows 32 bit has a limit to how much memory can be addressed, which affects all applications.
- Optional: If you want to analyze the raw data from usaspending.gov yourself, you will need the Java JAR file from www.usgcontractors.info. You will have to use a Windows machine with Java installed. For information on how to install Java on your Windows machine, look here: https://www.java.com/en/download/help/download_options.xml#windows.
- You can explore the .csv data, identify and select networks, and output the results for Netdraw manually in a text editor or spreadsheet. However, I have

written some scripts in STATA to facilitate that process. To use them, you must have STATA installed. STATA may be purchased here: <http://www.stata.com/order/>. Note the special academic rates. (Scripts for R will be available soon.)

- You may analyze and visualize the resulting networks using Netdraw by Steve Borgatti. You may freely download and install a copy of Netdraw here: <https://sites.google.com/site/netdrawsoftware/home>.
- Alternately, you may want to send the network data to Pajek for visualization and analysis. You may freely download and install Pajek here: <http://pajek.imfm.si/doku.php?id=download>

b. Analyzing raw data (if you need to)

If you have chosen to analyze the raw data yourself, go to <http://www.usaspending.gov/data?carryfilters=on> and use their web form to select the contract data that you are interested in, by time frame or agency. At present, only the information on contracts contains the fields necessary to identify networks, although this may change in the future. Accordingly, for now, under “Type of Spending” select “Contracts.” Under “Level of Details” select “Basic.” Under “Select Format” select “CSV.”

The file will be produced and be ready to download in zipped form. Download your file, then unzip it into your working directory.

If you requested a very large data file, it may take some time for the file to be available to download. You will be given a link that you can use to check on the status of your download.

You may have memory problems. If the file is too large, you may not be able to open it or run CNES on the data successfully. A year’s worth of federal procurement data may be too big for the 32 bit Windows operating system. If you can’t open the file in Excel - if Excel hangs - the file is too big and you should use a 64 bit Windows system. At present, there is no fix for this in CNES. (You could split the file using a file splitter, but this would defeat the purpose of identifying networks across the entire data file.)

Run CNES

Run CNES from the command line with the command “java -jar cnes.jar “filename” [-n]. The argument “filename” should be replaced by a filename ending in .csv without quotes. The brackets around “-n” indicate that is the optional argument for turning on name-based matching. CNES will look in the current directory for your data file unless you provide the full path. It expects your file to be in csv format.

It takes about 30-90 minutes for CNES.jar to crunch a year's worth of federal procurement data on my Intel 7 Windows machine with 16 GB of RAM while it is running in the background and I am doing other things. Your mileage may vary.

c. Explore Networks

Run STATA. Load the CNES5 utilities with the "do" command followed by the cnes5.do file with the full pathname.

Example: do "C:\Stata\mySTATAprograms\cnes5_util.do"

Then load the csv data file that you either downloaded from the www.usgcontractors.info website or generated with cnes.jar. To do this, give the STATA command:

cnes "filename"

where "filename" is the filename and full path of the .csv file to be analyzed. STATA will load it in and calculate the number of members in each network. You can then use STATA, and the included utilities, to identify networks of interest and to export all data in memory to either VNA (Netdraw) or Pajek.

3. Data Dictionary

The file that results from this analysis contains a list of observations. Each observation is a directed edge, containing information about a parent and about a child, and about the network in which they both may belong. The variables are:

Variable name	Description
contractorname	Derived from the vendorname, the standardized name has any leading or trailing blanks trimmed off and standardizes references to corporate forms
duns	Derived from dunsnumber, usaspending.gov, the DUNS number of the contracting company. The DUNS number is standardized to nine digits by trimming off the four right hand digits if the number is greater than nine digits
parentname	Derived from mod_parent, usaspending.gov, the standardized name has any leading or trailing blanks trimmed off and standardizes references to corporate forms

parentduns	<p>Derived from parentdunsnumber, usaspending.gov, the DUNS number of the parent company, if provided. The DUNS number is standardized to nine digits by trimming off the four right hand digits if the number is greater than nine digits.</p> <p>Or, if a name was provided but no DUNS number and the name-based matching option was selected, this field will contain:</p> <ul style="list-style-type: none"> • The DUNS number of the first business found that has the same name and a listed DUNS number. Or, if there is no existing name match with a DUNS number, • A generated identifier that is given to all companies of the same name. Generated identifiers can be distinguished from DUNS numbers because they are negative numbers.
network	An identifier that is given to all edges in the same network
nummembers	The number of edges in the network to which this arc belongs
matchedbyname	Set to "1" if the parentdunsnumber was derived from name based matching
signeddate	This is from signeddate in usaspending.gov, it provides a date on which this parent-child relationship was reported. Where more than one parent is reported for a contractor, it would allow you to see which is the most current.

4. Using STATA to Further Analyze Your Data or Export It

You can use STATA to explore your network data further. There are several CNES utilities to help you explore. You can:

- drop if nummembers<X. This will let you drop all networks that have less than X members. You may want to drop all networks that have only one member as being uninteresting, but if you are looking for a particular network you should keep these single member networks.

- `tab nummembers`. This will show you the distribution of components by the number of members.
- `list network if nummembers==X`. See which components are of size X.
- `listnet int1 int2 int3 . . .` List the members of the networks that you specify by network number.
- `listnetr regex`. List the members of any components whose members contain the string specified by the regular expression. Note that the regular expression can contain wildcard characters; if such special characters are to be interpreted literally, they must be preceded by a backslash. For more on regular expressions, see <http://www.stata.com/support/faqs/data/regex.html>.
- `findnet regex`. List the component numbers of any components whose members contain the string specified by the regular expression. Note that the regular expression can contain wildcard characters; if such special characters are to be interpreted literally, they must be preceded by a backslash. For more on regular expressions, see <http://www.stata.com/support/faqs/data/regex.html>.
- `stata2vna` - modified from the `stata2pajek` module written by Gabriel Rossman. See the original here: <http://ideas.repec.org/c/boc/bocode/s457075.html>. This command outputs all data in STATA's memory as a VNA file, so drop any networks that are not of interest. This works best if you have just one network in memory.
- `stata2pajek` - modified from the `stata2pajek` module written by Gabriel Rossman. See the original here: <http://ideas.repec.org/c/boc/bocode/s457075.html>. This command outputs all data in STATA's memory as a VNA file, so drop any networks that are not of interest. This works best if you have just one network in memory.

Note: The tools `findnet` and `listnetr` are important because networks are often fragmented in the `usaspending` data. A single company may show up in more than one component - which means that those components are actually fragments of a larger network. This can happen because the company may have multiple DUNS numbers and the linking entity is not part of the `usaspending.gov` database.

5. Visualizing in Netdraw

When visualized in Netdraw, the nodes are the DUNS entities (locations of a business or co-located legal subdivisions), and arrows are drawn from the possible parent company to the child. The variable `matchbyname` is also included; you can set the lines between nodes to a different color to see where connections were made with name matching (less sure). Because the current version of CNES forces one consistent name on all DUNS numbers, there will be few matches by name.

6. Known data problems - many!

There are numerous known data problems, which is why this software outputs **possible** contractor networks rather than actual contractor corporate networks. ***It is up to you to confirm the data before using or publishing it!***

First, the files downloaded from usaspending.gov may contain incorrect data. They may list a parent company incorrectly or not at all. They may not have the correct DUNS number associated with the companies. The information may be out of date, failing to reflect transfers of business ownership or other structural changes. Business names may be incorrectly or inconsistently entered.

Second, the only data in the usaspending.gov database is about government contractors and their parents. If there are parts of the network that do not contract with the U.S. government, they will not show up as members of a network, or the network may appear to be multiple networks instead of one.

Third, business entities are identified using their Dun & Bradstreet DUNS number. While in theory every business location and collocated legal subdivision of a business gets a unique DUNS number, in practice, this may not be the case. For example, a business that is sold would get new DUNS numbers, but may continue to use the old ones.

Fourth, if the CNES is used to do name-based matching, this carries the risk that two business entities with the same name will be incorrectly conflated. This is particularly likely to happen with legal subdivisions and locations, which often have similar or identical names; with less frequency, it may also happen with wholly unrelated companies. On the other hand, because company names are not entered consistently, name based matching will show a company whose name has been entered slightly differently in two different observations as two different entities when in reality it is the same entity.

Note also that where multiple names are entered for the same DUNS number, CNES picks one at random and enforces that name across all entries for that DUNS number.

Unfortunately, resources do not permit user technical support. However, if you find a bug or have a suggestion for expanded functionality, please send it to **bugs@usgcontrators.info**.