User Guide for the Patent Network Visualization Tool

Purpose

The purpose of the Patent Network Visualization Tool is to map out the social network of patent inventors. There are a number of reasons this could be useful:

- You might want to know which companies were major players in a technological field, in an intuitive, visual way.
- You might start with an inventor and want to know who she collaborated with at an important juncture in her career. Was she a loner? Immersed in a dense web of inventors working on similar topics? At the periphery of the field, or a key figure connecting several networks?
- You might be curious about ways that a field (or inventor's career) changed over time. Comparing several social networks representing different periods can visualize these changes.

The goal is to be able to see *around* a given technological field in an intuitive way, quickly and easily.

Getting Started

The simplest option is simply to press "Generate," then look in the right-hand "Output" column for links to visualizations. This will use default values.

If you want to change these default values, you can either manually revise them, or use the optional Inventor Name or Patent Sub-class searches to find new inventors to serve as origin IDs.

Inventor Name search will look up the inventor IDs from our database for any given name. This will erase the default values the first time you do it. The second, third, etc. times you use the search, it will simply add the new search results onto the old ones.

Inventors will likely have multiple IDs, because our algorithm tries to keep track of separate inventors with identical names. If it looks like this search has found the wrong person, simply remove the unwanted inventor ID from the comma-separated values box before pressing 'Generate.'

Patent Sub-Class Search finds everyone patenting in a particular technology (within the chosen dates), using US Patent and Trademark Office technology classifications. These can be found at http://www.uspto.gov/web/patents/classification/ or via the Classification section of a Google Patent page.

Both of these options are described in more detail below.

The Data

The data underlying this visualization is the Fung Institute for Engineer Leadership's patent database, which draws and processes data from the United States Patent and Trademark Office (USPTO).

The data covers 1976 to the present. Unfortunately, before that year the USPTO did not require the same type of data from patent applicants, so this is the current standard. Google Patents goes back further, but even its earlier data relies upon scanned documents and is much less usable further back in time.

The Fung Institute dataset has been processed and disambiguated, so in theory, each individual inventor should have one unique ID assigned. This means you can differentiate John Smith from Wisconsin and John Smith from Arkansas as two separate people, but still know that John Smith from Wisconsin is the same person after he moves to Nebraska. In practice, this process is still imperfect, but it nevertheless makes this data set uniquely powerful at tracking the movement of inventors.

A full, technical description of the data can be found at this following page: Description of Data

Bulk downloads of this data for research purposes are also possible: Bulk Downloads

Note: Ongoing Data Issue. As of March 30, 2015, we have discovered that some patents will not show up in these visualizations because their application dates did not import correctly into our database. Until we can re-import these dates from USPTO data and check that it is complete (estimated: Summer 2015). A first-guess estimate is that ~5% of patents between ~1995-2005 are missing this information, though hopefully that is an overestimate. This means that some links that should exist between inventors do not show up. So, diagrams should be taken as "at least this many links existed," but possibly some are missed until we can fix this underlying data issue.

The Social Network Diagrams

Interpreting the Network Diagrams

The size of the dots representing an inventor are proportional to the number of "future cites" the inventor has received. That is, if another patent ever cites this inventor's patent as a prior art inspiration, the inventor receives a "future cite." Thus, an inventor who had 10 patents, each of them cited by 10 different future patents, will be the same size as an inventor with one patent cited by 100 others. By far, the most common number of future cites a patent will receive is 0. The color of the dots is determined by the assignee on his **most recent** patent (within the chosen dates).

Note on patent assignees

For those less familiar with patenting, inventors do not always (or even usually) receive the legal control over a patent. When an inventor working for a company or a university files for a patent, the company/university usually gets "assigned" the patent. The inventor's name remains on the patent under "Inventors," but this is primarily for purposes of informal credit rather than legal meaning.

Thus, the most recent assignee on a patent likely represents the company/university the inventors worked for when filing it. This is an imperfect system, as inventors can (and do) switch jobs, but is still a useful shortcut.

The 10 assignees represented most often in the final data are automatically assigned different colors. Each inventor's dot is then assigned the color of this most recent assignee, as a proxy for affiliation.

How it Works

The tool, in theory, can map any number of layers of co-inventors' co-inventors' co-inventors' co-inventors' .... In practice, we limit to 3 generations, as described below, because the results get unwieldy quickly.

We start from an initial set of inventors, using the unique inventor IDs in our database. (In practice, each inventor may have multiple IDs, since names are not always uniform in the database, despite efforts to disambiguate the results).
We will call these people the "seed inventors."

When you first load the page, there are default seed inventor IDs already entered.

One "generation" – co-inventors:

The tool finds all patents filed by the seed inventors within the time frame entered. For example, John Doe might have filed for patents 1002043 ("Better mousetrap") and 43421110 ("Best mousetrap") during this window. If the inventors listed on "Best moustrap" are John Doe, Thomas Jefferson, and Albert Einstein, the network will create three connections:

John Doe: Thomas Jefferson
John Doe: Albert Einstein
Thomas Jefferson: Albert Einstein

It repeats this process for each patent. Note that it does not (at this point) care whether Jefferson and Einstein co-invented on one patent or one hundred.

The following diagram consists of 7 patents filed by a total of 13 inventors:
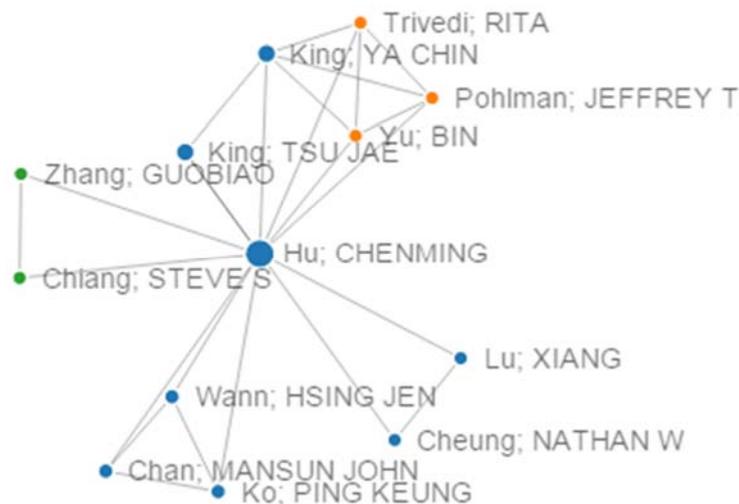
You can see everyone with whom our seed inventor (Chenming Hu) filed a patent as a co-inventor in this window (1998-2000), and also the relationships among these other patentees. On one patent, for example, the inventors were Chenming Hu, Nathan W Cheung, and Xiang Lu.

Two "generations" – co-co-inventors:

At two generations, we take things a step further. Now the tool treats all of the names it found in Step 1 above (one "generation") as seed inventors, and runs through the process again. It finds all of the patents they applied for in the chosen time frame, then finds all inventors affiliated with each of those patents.
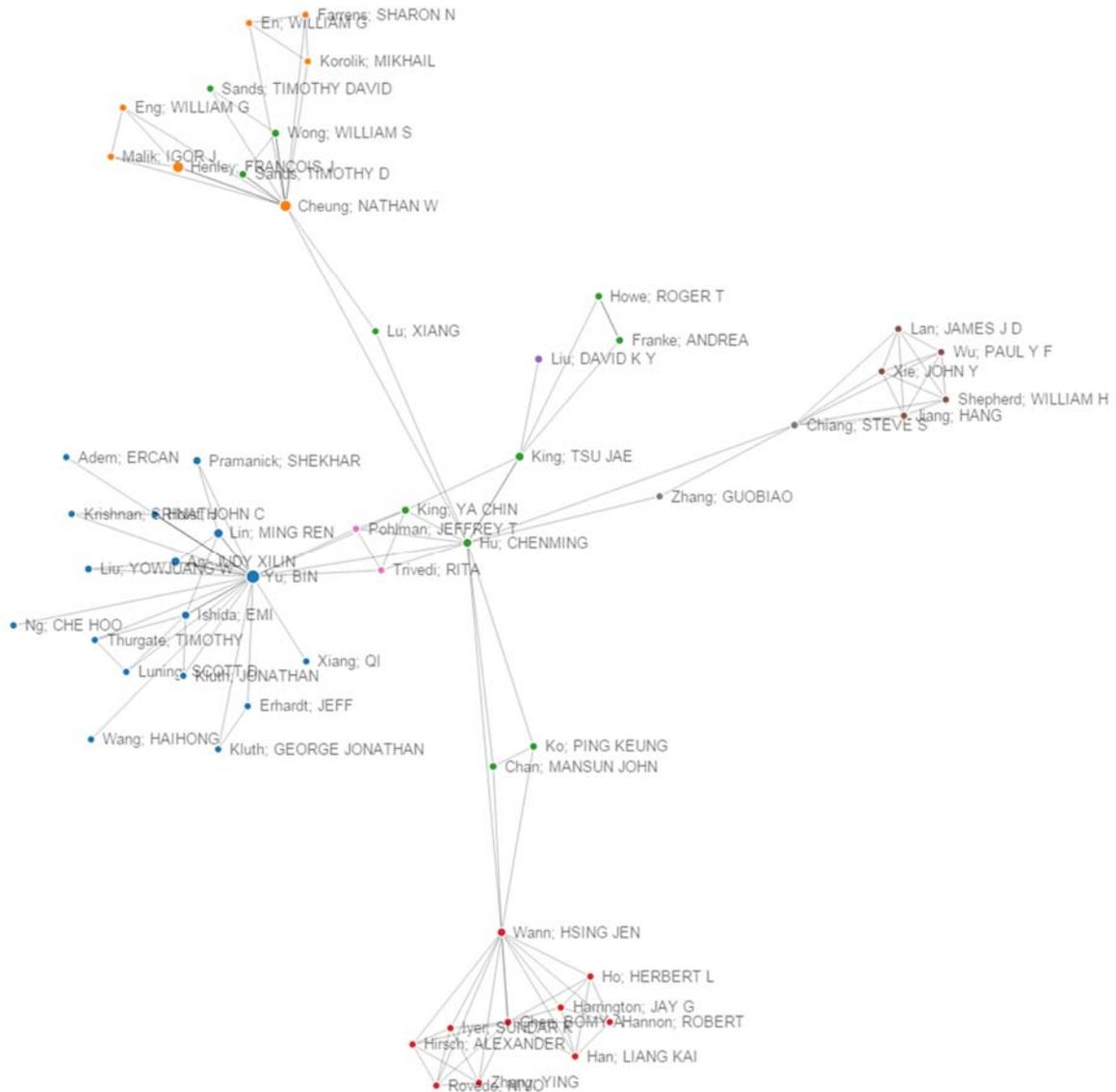

The initial picture is still the same, but expanded out another degree of co-inventorship.

We're beginning to see that there were entire networks of co-inventors out there, but that they have very little in common – other than their affiliation with Chenming Hu, our seed inventor.

Three "generations" – co-co-co-inventors:

By the time we get to three generations, the number of inventors being mapped can get very large indeed.
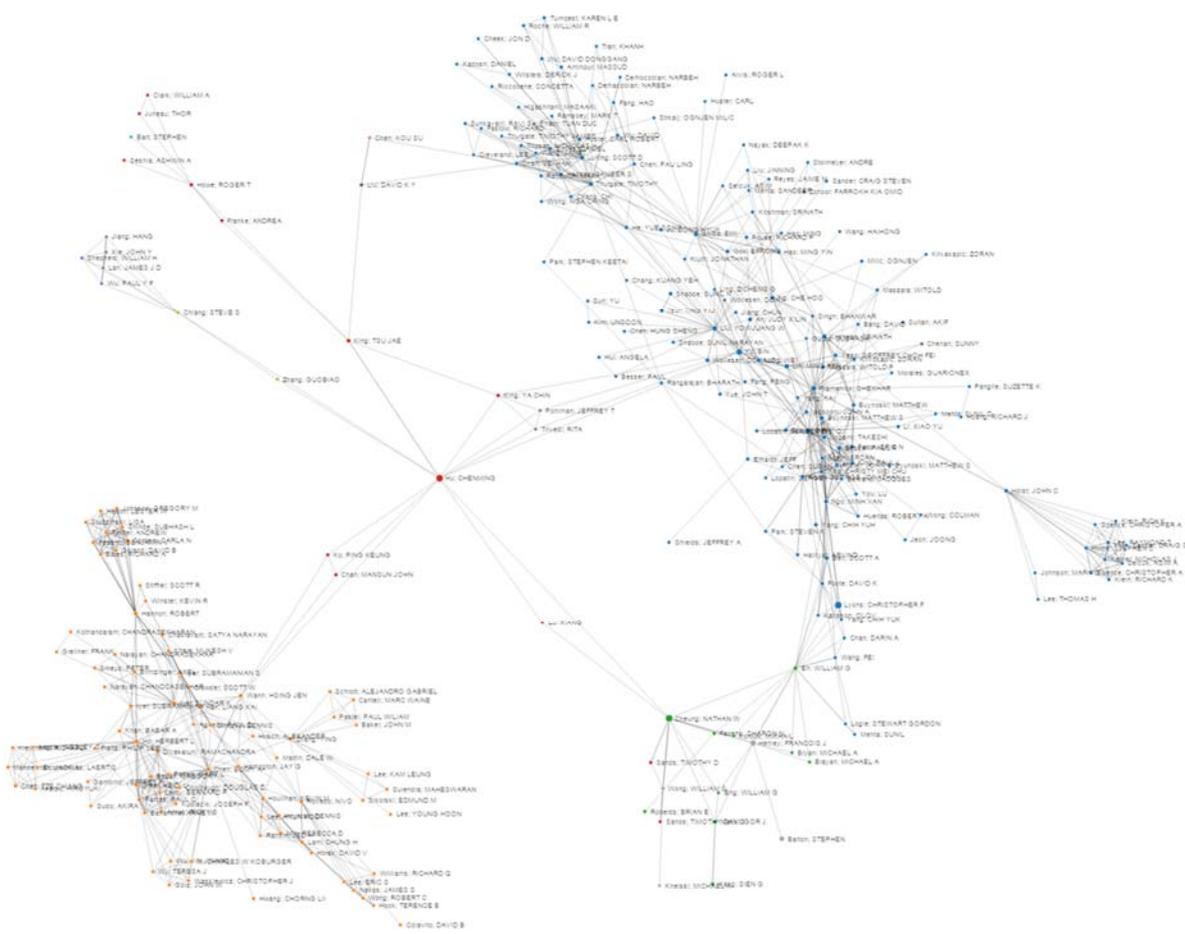
We now see more clearly than ever that our seed inventor, Chenming Hu, was truly a knowledge broker, connecting various communities in a way difficult to make clear through interview or archival work alone.

Using the Tool

Option 1: Use the Defaults / Manual entry

When you first load the tool, it is ready to generate a map using some pre-set, default values. That is why at first, Chenming Hu, Tsu Jae King Liu, and Jeffery Bokor show up in the Inventor IDs section.

You can erase these default inventor IDs as desired, change the dates, or adjust the number of generations to iterate.

Simply click "generate" to create a social network diagram for these settings. Once you see the success message ("**Success! Please follow [this link](#) to your patent network map.**"), click the link to render the image.

Option 2: Search by Inventor Name

Ignore the default settings if you wish to create a map for a new inventor (or inventors). Simply enter a first name and last name into the "Search by Inventor Name" box and press 'Search.'

The first time you search for a name, it will replace the defaults. Any subsequent names you search for will be ADDED to earlier results. Thus, if you want to create a network diagram using three inventors as the seed inventors, simply search for them one after another.

When you search for a name, the tool will do its best to find all possible variations on the name (ie, both Chen Ming and Chenming). If one of the names it finds seems incorrect, just delete the corresponding Inventor ID from that box before you generate the social network.

Once you're ready, press "Generate" and you're set!

Option 3: Search by Patent Sub-class

Possibly you're interested in a really big-picture view of an industry in a given time frame. This option is an attempt to balance between big-picture and reasonable stress on the system.

As part of the process of granting a patent, the USPTO assigned a number of patent classifications and sub-classification. For example, patent classification 438 marks patents in the technology area of "Semiconductor Device Manufacturing: Process," and there are almost 1000 sub-classifications for more specific technologies within this broader description (for instance, "Temporary Protective Layer" or "Specified etch Stop Material").

7

You can find the patent classification assigned to a patent on its Google Patent page. For example:

https://www.google.com/patents/US6413802

The number of patents filed per year in a given sub-class will vary widely across technologies, but might be as high as hundreds or thousands.

In order to use this option,
FIRST, enter the DATES you intend to search. Because these results get very big, very fast, please choose a narrow window (~1-2 years is suitable, 3-4 gets more difficult to generate, render, and interpret the results). The tool will use these dates in searching for seed inventors related to this patent class.

SECOND, enter the chosen main class and sub class into the appropriate boxes and press 'Search.' The tool will now find everyone who applied for a patent within the time window selected, for which the patent eventually received the appropriate classification. These inventors will now be the seed inventors.

As you will notice, this usually generates many seed inventors, sometimes hundreds even for a given year. This is why we recommend only 1-2 generations.

Images generated from this option will give images something like the following, which maps the semiconductor industry from 1998-2000 (2 generations):
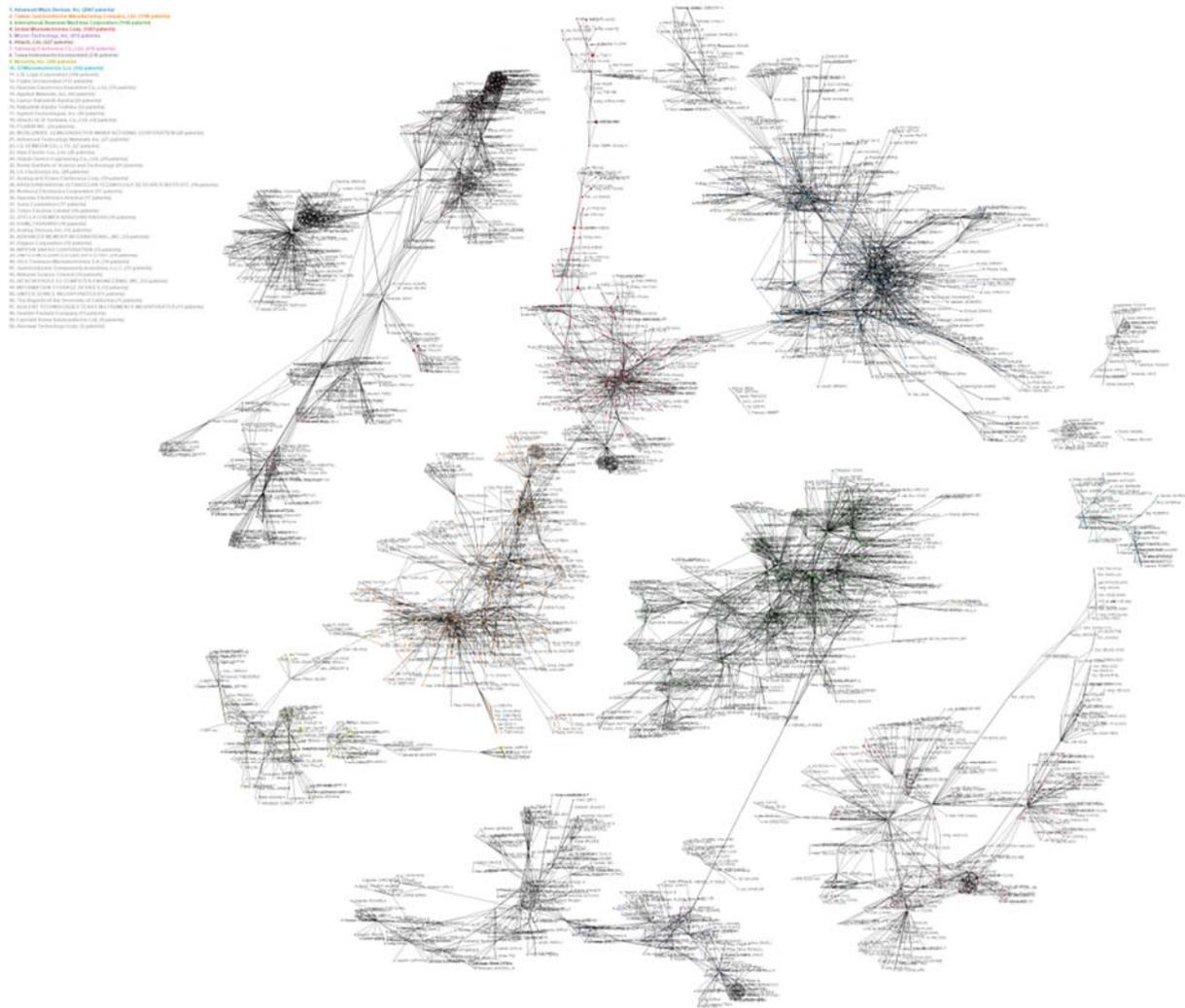
Rendering and Manipulating the Social Network

The network treats each inventor like a charged particle, repelling all others to which it is not linked. That means that it can take some time for the network to sort itself out into a coherent picture.

If it seems that the image is unnecessarily complex, you can "untangle" the results manually. To do this, simply use your mouse to drag a specific inventor in the direction of your choice. Connected inventors will be drawn along.