



[GOOGLE FUSION TABLES] NETWORK GRAPH

Tutorial by Iman Salehian (UCLA) & David Kim (UCLA)

Google Fusion Tables is a Google Drive-based application that allows for the creation and management of spreadsheets, making data visualization the ultimate end of this collaborative workflow. While it offers a bevy of visualization options, ranging from constructing basic pie charts to mapping tables of coordinates, this tutorial focuses on its Network Graph capability, a feature that allows for network visualization and analysis.

Why Google Spreadsheets and Fusion Tables?

As users of Microsoft Excel will notice, Google Fusion Tables is essentially a spreadsheet with rows and columns to create or import data. It is a web-based application that allows the users to collaboratively create/edit the spreadsheets remotely. Google Fusion Tables also offers Network Graphing, a simple tool that familiarizes users with the rudiments of network visualization. Depending on the complexity of their data and desired visualizations, some may find Google's Network Graph capability to satisfy their needs, while others may consider it a useful stepping-stone to more customizable visualization software such as Gephi or Cytoscape.

Getting Started

While Network Graph works with any .csv file, as beginners, it behooves us to start from scratch in order to familiarize ourselves with the back-end workings of this visualization tool. The following assignment will walk you through the process of visualizing a network from a data table you will construct, while posing a series of questions encouraging students to consider overarching data visualization concepts.

NOTE: This tutorial is tailored for those users with access to Gmail and a Google Drive account. If you don't have a Gmail account, you may use Excel or an alternative spreadsheet creator to get started, but are encouraged to create an account so as to be able to easily save and access your work.

Part One: Data

1. Collecting data

Naturally, data visualization needs data; it follows that network visualization needs networks. What, then, is a network? At its most basic level, a network can be defined as a group of objects or entities—referred to as “nodes”—linked by relationships—referred to as “edges”. These visualizations are useful in representing a veritable slew of relationships, capable of representing links between employees and companies, pets and owners, friends and more friends, and so on.

For this assignment, feel free to use any network you'd like, so long as you can identify consistent relationship and object types within that group. As you will be asked to compile a list of 40-50 relationships within a group of objects/entities, aim to document an accordingly rich network. For the purposes of this tutorial, we will examine the complex network maintained by the character's of television drama, *Lost*.

- a. Take a moment to **create a list of relevant objects/entities**.
 - i. In our *Lost* example, we will consider the main characters of the show.
e.g. *Jack Shepard, John Locke, Ben Linus...*
- b. Next, aim to **categorize your objects or entities**. Do any “types” come up?
 - i. The characters in *Lost*, for instance, are often defined by their membership to parent groups. We will use the labels: “Others,” “Tailies” and “Core Group,” established in seasons one and two.

Does your network consist of “heroes” and “villains”? “Family members”? “Friends”? Considering the small size of the data sample you are creating, ideally **create two (2) and no more than four (4) categories**.

When filling in your spreadsheet, include the labels you decide upon in parenthesis next to the object/entity they describe.

e.g. *Person A (Villain) or Jack Shepard (Core Group)*

- c. Take a moment to **identify consistently emerging relationship types** and decide on the best description to use for these relationships: friends, enemies, married, etc. **Aim to document 3-5 consistently appearing relationship types**.
 - i. For our *Lost* example, we will use the generic relations listed below.
e.g. “*Friends with,*” “*Foes with,*” “*Family with,*” and “*Romance with*”
 - ii. You’ll notice that these relationships are all mutual. This is characteristic of an *undirected graph*, a network graph in which the relationships--as its title suggests-- have no specific direction. Relationships that are unreciprocated (e.g. *Parent to child, Teacher to student, Murderer to victim, etc.*) are featured in *directed graphs*, graphs that use arrows to describe the direction in which a relationship moves.
For now, limit yourself to describing reciprocal relationships.

Developing these consistent labels and relationship types will allow you to take full advantage of the search and filter features available on Google Fusion Tables. For instance, if you only want to see who is “*Friends with*” who or members of *Lost*’s “*Core Group,*” the search query can be used to filter for these specific qualities.

2. Creating a Spreadsheet

Proceed to populate your data table with the information you’ve collected, aiming to define 40-50 relationships between objects/entities. In this simple network visualization, your data table will consist of three columns: the first column featuring Object/Entity A, the third, Object/Entity B, and the second, the relationship the two maintain. Each row—excepting the first, which should list your column names—will, in essence, describe a relationship maintained within your network.

Here, the “types” we developed in 1.b will come in handy. List the category an object/entity pertains to in parenthesis besides his/its title, as is pictured below.

<i>Character A</i>	<i>Relation</i>	<i>Character B</i>
Jack Shepard (Core Group)	Foes with	Tom Friendly (Other)
Jack Shepard (Core Group)	Family with	Claire Littleton (Core Group)
Bernard Nadler (Tailies)	Family with	Rose Nadler (Core Group)

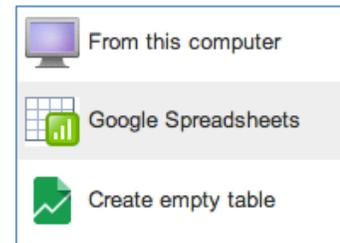
- a. Try to connect each object/entity with at least two other objects/entities. The more connections you draw, the tighter your network will appear.
 - i. For the purpose of this undirected graph, you do not have to repeat relationships that have been already established previously in the spreadsheet.

Part Two: Visualization

Once you have completed your spreadsheet, you are ready to plug it into Fusion.

1. Importing Data

- a. To begin, click “Create” under Google’s Fusion Tables app, found [here](#).
- b. If you did NOT use Google Drive’s Spreadsheet creator to create your table, go ahead and import your file in the “From this computer” tab. If you DID use Google to create your spreadsheet, click the Google Spreadsheets tab, and select the spreadsheet you created for visualization.



Step 1.b

- c. Review your spreadsheet to ensure it has imported properly and click “Next”.

Column names are in row

1	Character A	Relation	Character B
2	Jack Shepard (Core Group)	Foes with	Tom Friendly (Other)
3	Jack Shepard (Core Group)	Family with	Claire Littleton (Core Group)

Step 1.c

- d. Give your table a title and description. Check the “Export” box if you wish to make your data public and downloadable for future users.

Note: On occasion, the app may glitch and alert you that there were issues loading. Simply clicking "Finish" a second time usually resolves this issue.

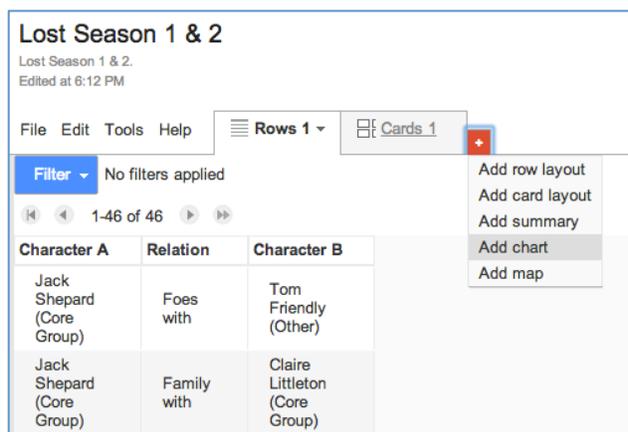
2. Visualizing Data

For step-by-step directions on Visualizing Data, follow Google's Tutorial on "Network Graph," [here](#), or follow its summary listed below.

a. A window will open featuring your data table. Beside the top row of tabs, you will find a small red square with a [+] sign. Click this and select, "Add Chart."

- Choose the "Network graph" option (visible at the bottom of the left side panel) if Google Fusion Tables

has not done so already.



The screenshot shows the Google Fusion Tables interface for a table titled "Lost Season 1 & 2". The table has three columns: "Character A", "Relation", and "Character B". The data rows are:

Character A	Relation	Character B
Jack Shepard (Core Group)	Foes with	Tom Friendly (Other)
Jack Shepard (Core Group)	Family with	Claire Littleton (Core Group)

The interface includes a menu with options: "Add row layout", "Add card layout", "Add summary", "Add chart" (highlighted), and "Add map".

Step 2.a

b. By default, the first two text columns will be selected as the source of nodes. Change these to whatever titles you have listed for your first and third columns. For the Lost example, they are Character A and Character B.

c. Considering the basic nature of our network graph, Network Graph's "Appearance" and "Weight" features are not of much use to us. Provided below are short descriptions of their hypothetical uses.

- For a run-down of what "Link is directional" implies, see this tutorial's description of the distinction between "directed" and "undirected" relations in 1.c.ii..
- "Color by Column" simply refers to coloring the nodes displayed according to the columns they pertain to.
- Weighting refers to assigning a value to your described relationships. At the most basic level, this would mean including a separate column in your spreadsheet that links numbers to relationships. The theoretical implications of such a task, however, tend to be a lot more

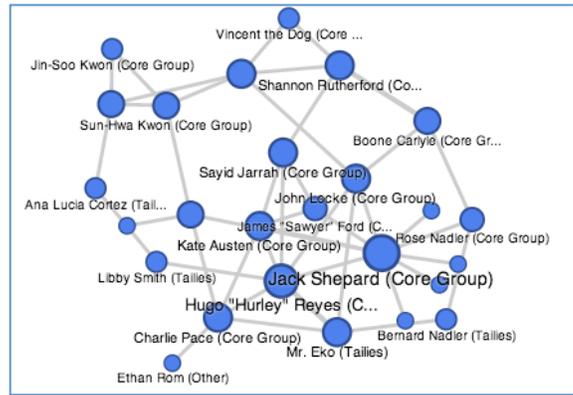
△ TROUBLESHOOTING

If at any point your graph is not appearing:

- Make sure you are displaying the full number of "nodes" available in the [___ of # Nodes] textbox. (INSERT PICTURE)
- Your filter options may be too refined for your small dataset. Removing a specification or two should solve the issue.

complicated, as they involve disambiguating subjective qualifiers such as relationship intensity or value.

- d. Your basic visualization should be complete and ready for searching and filtering!
- Click and drag to navigate within the network graph window. You may also click and drag specific 'nodes' to rearrange your network graph.



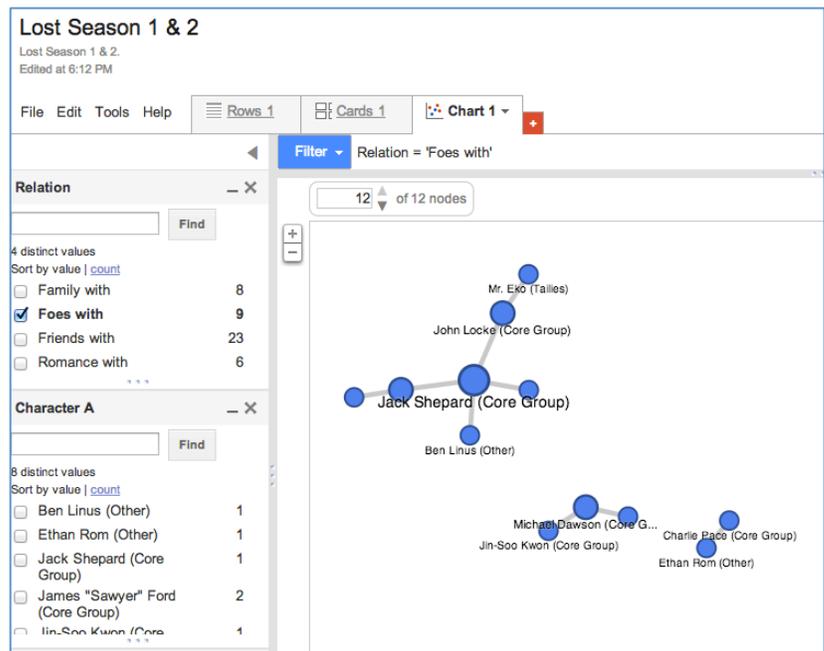
Step 2.d

3. Search/Filter

So, you've completed your first network visualization. Now what? An added benefit of having visualization online lies in our ability to interact with and filter it.

- a. In the top left corner of your window, you should see a blue "Filter" button. Click "Filter" and choose a character list (i.e. column of your original spreadsheet) you wish to filter. You may wish to select all three, as all will remain as menus on the left hand panel.

- b. Check the boxes of specific relations or objects/entities you wish to filter for, or use the search boxes to specify object/entity 'types'.



Step 2.a, 2.b

e.g. For our *Lost* example, we may input the search term "(Core Group)" to see only those relationships maintained by the "Core Group".

Remember! You must input this search term in BOTH your first and third column/character list/etc. to filter out all excess object/entity types.

You have successfully graphed and filtered your very own network graph!

Part Three: Challenges in Visualization

Comparing your final Network Graph to the information rich spreadsheet you created in Part One, you may understandably find yourself frustrated with the limited information being represented. This may mean it is time to move on to a more sophisticated visualization tool, such as Gephi or Cytoscape.

Don't, however, let appearances fool you. While ostensibly simple, Network Graph envelops its own set of theoretical challenges.

The "Frenemy" Complex: The Limits of Labeling

In almost any visualization project, one is invariably forced to ask oneself, is it really this simple? Am I misrepresenting anything? This monster of complexity probably reared its head the moment we set out to "define" something as volatile as a relationship between individuals.

Within our *Lost* example, we aimed to limit the complexity of the network by using a smaller set of data (i.e. Seasons One and Two), but as fans of the show will note, even these episodes are rich with challenges. Protagonists, Jack and Kate, for example, maintain a relationship that shifts between "Romance with" and "Friends with" from episode to episode.

Within your dataset, ask yourself: Are relationships always 'reciprocated'? (directional?)

Could a connection between two entities be more than one?

e.g. A and B can be friends, but 'secretly' B might also consider A to be an enemy, aka frenemy.

How would you visualize these more complex, ambiguous, and one-directional type of relationships?

Furthermore, what happens when objects/entities defy a single label? Within *Lost*, for example, many of the 'parent' groups we identified our characters with either unite or further divide, complicating the superficial labels we applied to the characters that comprise these groups. A "Tailie," for instance, can be said to be absorbed by the "Core group." Within your data set, are there entities that belong in more than one types? Would you assign more than one type ('tags') to an entity? And, ultimately, what challenges do you perceive in "disambiguating" an entity's types and relationships, when their "real life" counterparts prove more complex than a single line of description could ever hope to convey?

Part Four: Advanced Topics

Relationship Index

As the battery of questions above may lead you to realize, specificity is oftentimes a must when working with ambiguous or subjective data. If you are planning on making Network Visualization central to your study of a particular topic, consider creating an "index" for relationships that defines the terms you are employing within your spreadsheet and, by extension, your graph. Define what conditions/qualities are invoked by the term "friend,"

"enemy," and so on.

Spatial/Temporal Dimensions

While beyond the scope of Network Graphs, keep in mind that work is being done in the field of adding spatial and/or temporal dimensions to network graphing. While these functions remain beyond the scope of Fusion at this point in time, consider the implications of creating a column for GIS data and layering a visualization over Google maps. How would a network graph be enhanced by adding a time stamp or period?

Users will find that these hypotheticals become a reality with a more advanced graphing counterpart, Gephi.