

DATA PREPARATION FOR GEPHI: STEP BY STEP

What You Need to Accomplish

Here is what you have created from the short story Kung I-Chi: spreadsheet of character interactions (henceforth, the “Character Interaction” worksheet).

	A	B
1	Kung I-Chi	I
2	Tavern Keeper	Kung I-Chi
3	Tavern Keeper	I
4	Tavern Keeper	Customers
5	Kung I-Chi	Kids
6	Kung I-Chi	Mr. Ting
7	Kung I-Chi	Ho Family
8	Kung I-Chi	Customers
9	I	Customers
10		

For Gephi to read this data, you will need to transform it into two separate datasheets: a **“nodes” sheet** and an **“edges” sheet**. The “nodes” sheet will look like this. You must name the first column “Id,” but you may name the other columns anything you wish. You may also add as many columns as you wish. These are *attribute* columns, which describe the properties of each Id. All Ids must be unique.

	A	B
1	Id	Name
2		1 Kung I-Chi
3		2 Tavern Keeper
4		3 I
5		4 Customers
6		5 Kids
7		6 Mr. Ting
8		7 Ho Family

Here is the “edges” sheet. All edges sheet must have the following columns: **Source**, **Target**, **Type**, and **Weight**.

	A	B	C	D
1	Source	Target	Type	Weight
2	1	3	Undirected	1
3	2	1	Undirected	1
4	2	3	Undirected	1
5	2	4	Undirected	1
6	1	5	Undirected	1
7	1	6	Undirected	1
8	1	7	Undirected	1
9	1	4	Undirected	1
10	3	4	Undirected	1

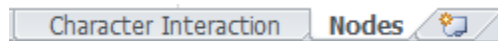
How To Do It?

The data transformation consists of two stages. In the first stage, you will need to create a nodes sheet, where each node is assigned a unique Id. In the second stage, you will create an edges sheet, where all relations between nodes are expressed as relations between Ids.

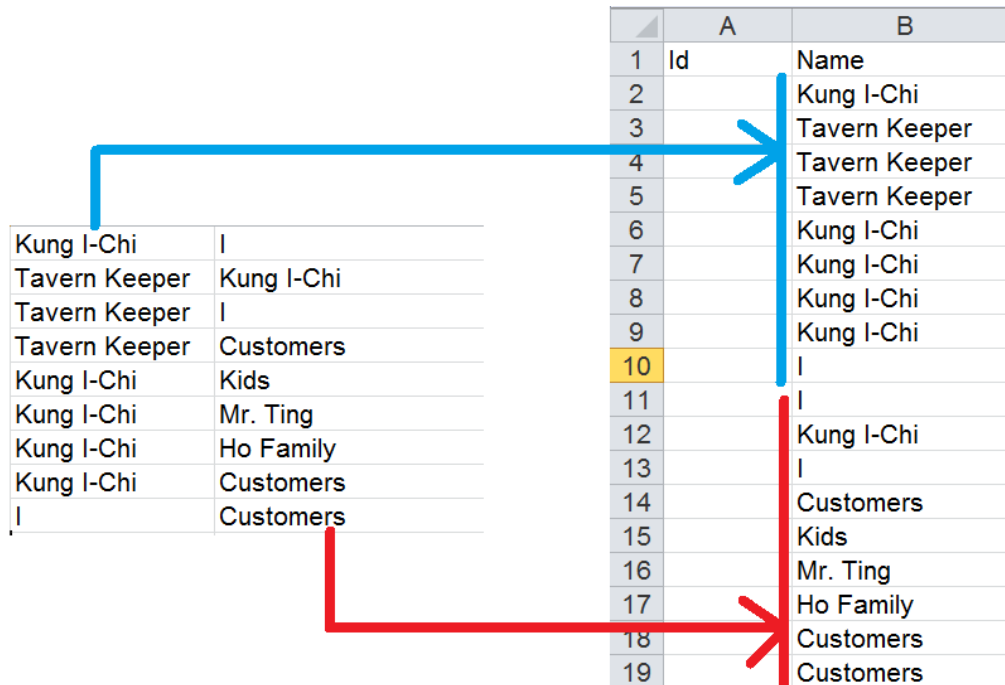
Before you start, you need to make sure that each character in your “Character Interaction” sheet has a unique name and this unique name is used consistently in each reference to the character in your sheet.

Stage 1: Create the nodes sheet

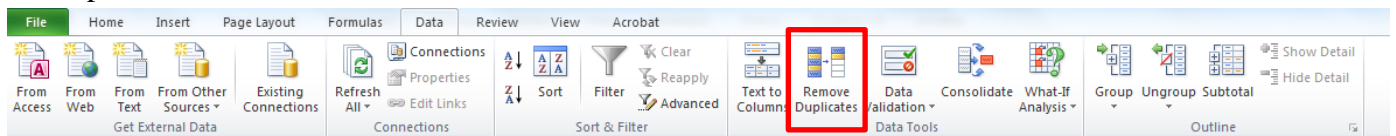
1. Create a new worksheet in Excel and name it “Nodes” (or anything of your choice).



2. Copy & paste all characters from the “Character Interaction” sheet to a *single column* in the Nodes sheet you have just created. This gives you a list of all characters in the short story.

A diagram illustrates the data transfer process. On the left, a table represents the 'Character Interaction' sheet with two columns: 'Name' and 'Id'. The 'Name' column contains: Kung I-Chi, Tavern Keeper, Tavern Keeper, Tavern Keeper, Kung I-Chi, Kung I-Chi, Kung I-Chi, Kung I-Chi, |, Kung I-Chi, Kung I-Chi, Kung I-Chi, Kung I-Chi, |, Kung I-Chi, Kung I-Chi, Kung I-Chi, |, Customers. The 'Id' column contains: |, Kung I-Chi, |, |, |, Kung I-Chi, Kung I-Chi, Kung I-Chi, Kung I-Chi, |, |, Kung I-Chi, |, |, Customers, Kids, Mr. Ting, Ho Family, Customers. A blue arrow points from the 'Name' column to column A of the 'Nodes' sheet. A red arrow points from the 'Id' column to column B of the 'Nodes' sheet. On the right, the 'Nodes' sheet is shown with columns A and B. Column A is labeled 'Id' and column B is labeled 'Name'. The 'Name' column contains: Kung I-Chi, Tavern Keeper, Tavern Keeper, Tavern Keeper, Kung I-Chi, Kung I-Chi, Kung I-Chi, Kung I-Chi, |, |, Kung I-Chi, |, |, Customers, Kids, Mr. Ting, Ho Family, Customers. Row 10 in the 'Nodes' sheet is highlighted in yellow.

3. Click any non-empty cell in the Nodes sheet. Then click Remove Duplicates to remove duplicate values in the “Name” column.



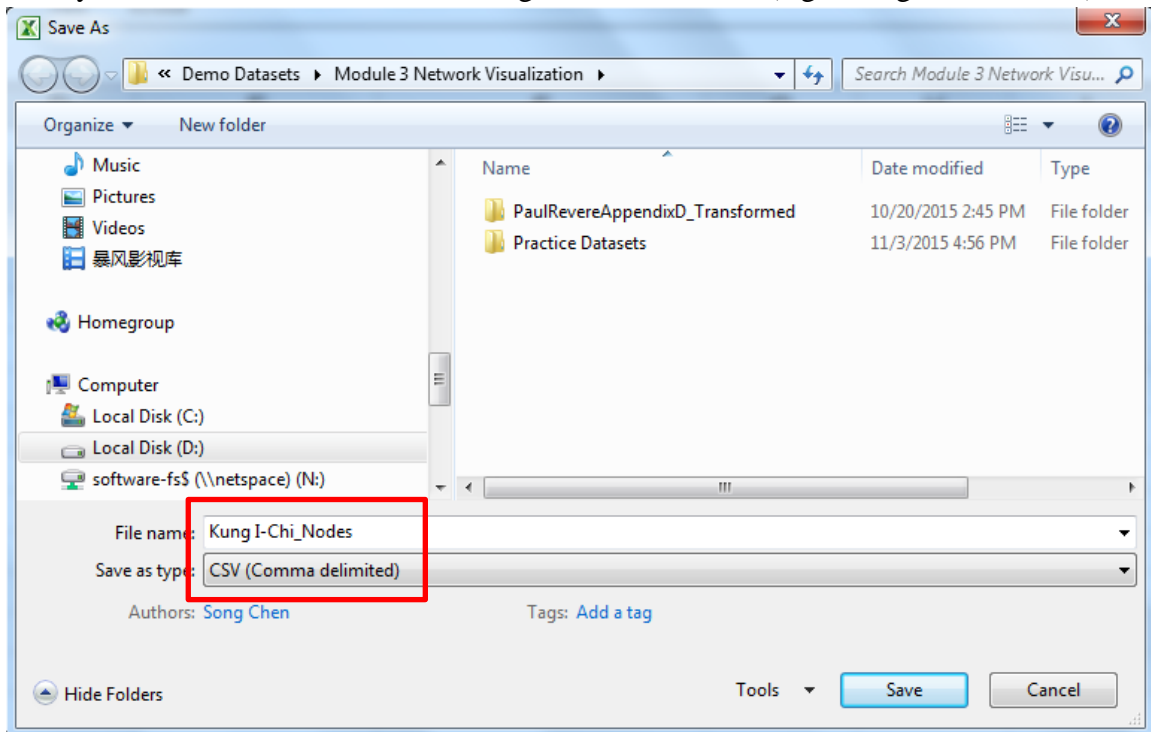
4. This gives you a list of *unique* characters in the short story.

	A	B
1	Id	Name
2		Kung I-Chi
3		Tavern Keeper
4		I
5		Customers
6		Kids
7		Mr. Ting
8		Ho Family

5. In the Id column, assign each character a unique Id.

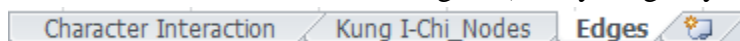
	A	B
1	Id	Name
2	1	Kung I-Chi
3	2	Tavern Keeper
4	3	I
5	4	Customers
6	5	Kids
7	6	Mr. Ting
8	7	Ho Family

6. Save your Nodes sheet as a **csv** file and give it a new name (e.g., Kung I-Chi_Nodes).



Stage 2: Create the edges sheet

1. Create a new worksheet in Excel and name it “Edges” (or anything of your choice).



2. Name the columns as follows.

	A	B	C	D
1	Source	Target	Type	Weight
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				

3. To prepare for the use of Vlookup function in the next step, copy the Ids to a column on the right of the Name column.

	A	B	C
1	Id	Name	
2		1;Kung I-Chi	1
3		2;Tavern Keeper	2
4		3;I	3
5		4;Customers	4
6		5;Kids	5
7		6;Mr. Ting	6
8		7;Ho Family	7

4. In the “Character Interaction” sheet, use the Vlookup function to convert all characters in the first column to Ids. For more on the Vlookup function, consult the following links:

<http://www.gcflernfree.org/excel-tips/how-to-use-excels-vlookup-function> ;
<https://support.office.com/en-us/article/VLOOKUP-function-0bbc8083-26fe-4963-8ab8-93a18ad188a1> ; <https://www.google.com/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#q=vlookup+excel&tbm=vid> .

	A	B	C
1	Kung I-Chi	I	=vlookup(A1
2	Tavern Keeper	Kung I-Chi	
3	Tavern Keeper	I	
4	Tavern Keeper	Customers	
5	Kung I-Chi	Kids	
6	Kung I-Chi	Mr. Ting	
7	Kung I-Chi	Ho Family	
8	Kung I-Chi	Customers	
9	I	Customers	

SUM =VLOOKUP(A1,Nodes!B:C

	A	B	C
1	Id	Name	
2		1;Kung I-Chi	1
3		2;Tavern Keeper	2
4		3;I	3
5		4;Customers	4
6		5;Kids	5
7		6;Mr. Ting	6
8		7;Ho Family	7

SUM =VLOOKUP(A1,Nodes!B:C,2,false)

	A	B	C	D	E	F
1	Id	Name				
2		1 Kung I-Chi	1			
3		2 Tavern Keeper	2			
4		3 I	3			
5		4 Customers	4			
6		5 Kids	5			
7		6 Mr. Ting	6			
8		7 Ho Family	7			

C1 =VLOOKUP(A1,Nodes!B:C,2,FALSE)

	A	B	C	D
1	Kung I-Chi	I	1	
2	Tavern Keeper	Kung I-Chi		
3	Tavern Keeper	I		
4	Tavern Keeper	Customers		
5	Kung I-Chi	Kids		
6	Kung I-Chi	Mr. Ting		
7	Kung I-Chi	Ho Family		
8	Kung I-Chi	Customers		
9	I	Customers		

	A	B	C	D
1	Kung I-Chi	I		1
2	Tavern Keeper	Kung I-Chi		2
3	Tavern Keeper	I		2
4	Tavern Keeper	Customers		2
5	Kung I-Chi	Kids		1
6	Kung I-Chi	Mr. Ting		1
7	Kung I-Chi	Ho Family		1
8	Kung I-Chi	Customers		1
9	I	Customers		3

5. Repeat the process to convert all characters in the second column to Ids.

D1		=VLOOKUP(B1,Nodes!B:C,2,FALSE)		
	A	B	C	D
1	Kung I-Chi	I	1	3
2	Tavern Keeper	Kung I-Chi	2	
3	Tavern Keeper	I	2	
4	Tavern Keeper	Customers	2	
5	Kung I-Chi	Kids	1	
6	Kung I-Chi	Mr. Ting	1	
7	Kung I-Chi	Ho Family	1	
8	Kung I-Chi	Customers	1	
9	I	Customers	3	

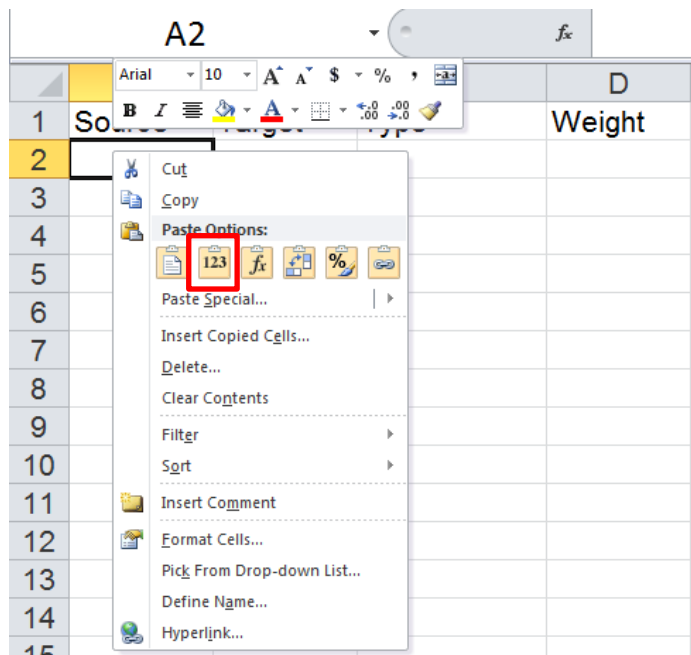
	A	B	C	D
1	Kung I-Chi	I	1	3
2	Tavern Keeper	Kung I-Chi	2	1
3	Tavern Keeper	I	2	3
4	Tavern Keeper	Customers	2	4
5	Kung I-Chi	Kids	1	5
6	Kung I-Chi	Mr. Ting	1	6
7	Kung I-Chi	Ho Family	1	7
8	Kung I-Chi	Customers	1	4
9	I	Customers	3	4

6. Copy the two columns with the Ids.

	A	B	C	D	E
1	Kung I-Chi	I	1	3	
2	Tavern Keeper	Kung I-Chi	2	1	
3	Tavern Keeper	I	2	3	
4	Tavern Keeper	Customers	2	4	
5	Kung I-Chi	Kids	1		
6	Kung I-Chi	Mr. Ting	1		
7	Kung I-Chi	Ho Family	1		
8	Kung I-Chi	Customers	1		
9	I	Customers	3		
10					

7. Paste them as values to the Edges sheet you have created earlier.

	A	B	C	D
1	Source	Target	Type	Weight
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				



8. Fill in the appropriate values for the Type and Weight columns.

	A	B	C	D
1	Source	Target	Type	Weight
2	1	3	undirected	1
3	2	1		
4	2	3		
5	2	4		
6	1	5		
7	1	6		
8	1	7		
9	1	4		
10	3	4		

	A	B	C	D
1	Source	Target	Type	Weight
2	1	3	undirected	1
3	2	1	undirected	1
4	2	3	undirected	1
5	2	4	undirected	1
6	1	5	undirected	1
7	1	6	undirected	1
8	1	7	undirected	1
9	1	4	undirected	1
10	3	4	undirected	1

7. Save your Edges sheet as a csv file and give it a new name (e.g., Kung I-Chi_Edges).

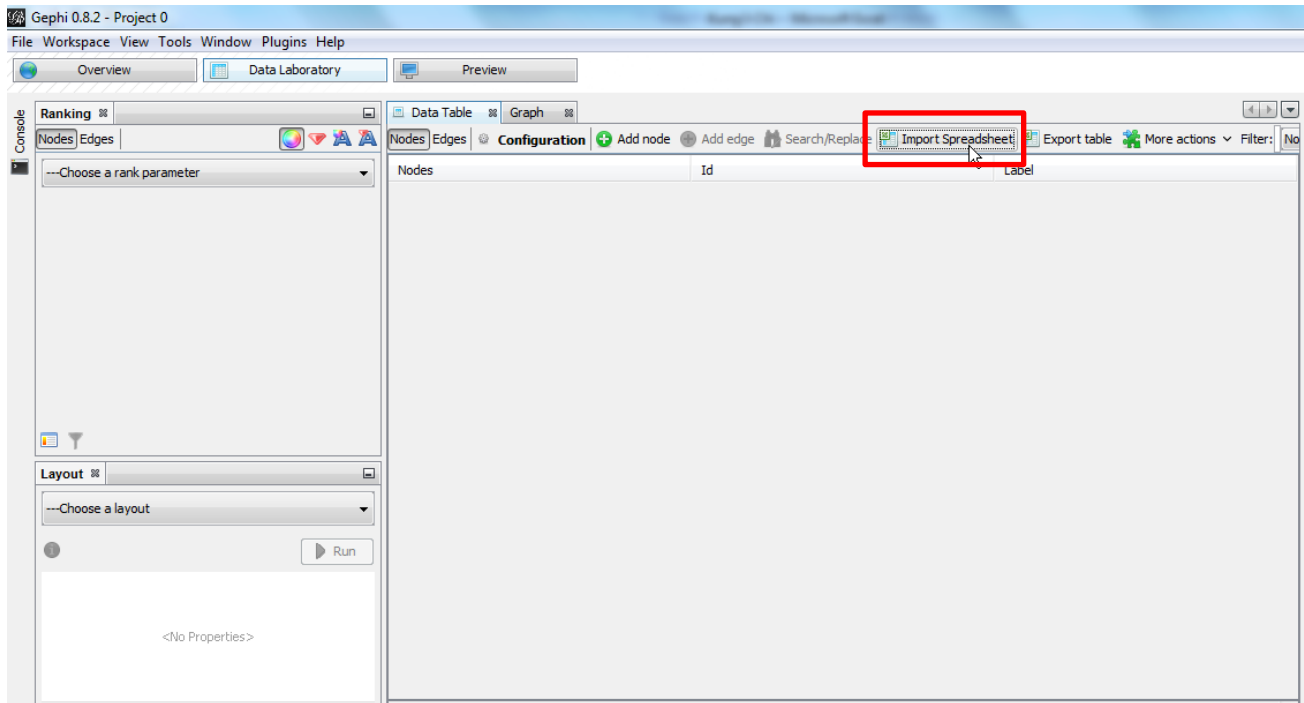
The screenshot shows a spreadsheet with the following data:

	A	B	C	D
1	Source	Target	Type	Weight
2	1	3	undirected	
3	2	1	undirected	
4	2	3	undirected	
5	2	4	undirected	
6	1	5	undirected	
7	1	6	undirected	
8	1	7	undirected	
9	1	4	undirected	
10	3	4	undirected	

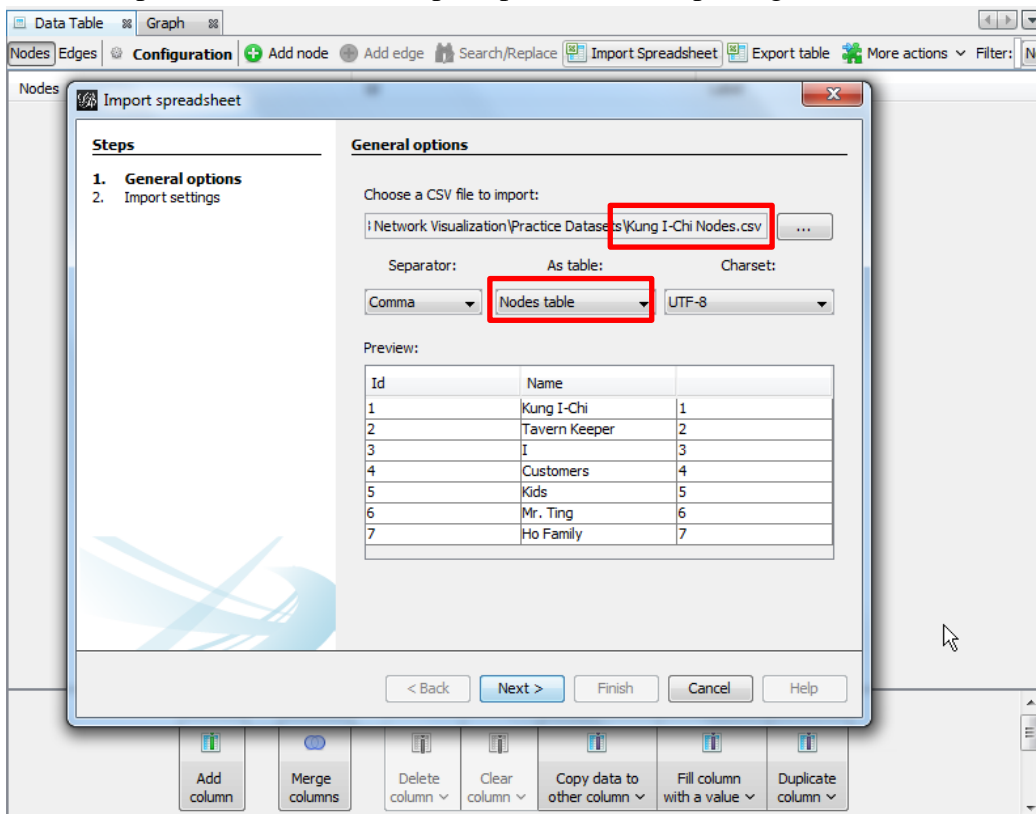
The 'Save As' dialog box is open, showing the file name 'Kung I-Chi_Edges' and the save type 'CSV (Comma delimited)'. The dialog box also shows the current directory path and a list of files in the folder.

Import the CSV Files into Gephi

1. Click “Data Laboratory.” Then click on “Import Spreadsheet” under the Data Table tab.



2. Choose the Kung I-Chi_Nodes.csv file to import. Make sure you choose “Nodes table” in the “As table” dropdown list. Follow the prompts to finish importing.



3. Choose the Kung I-Chi_Edges.csv file to import. Make sure you choose “Edges table” in the “As table” dropdown list. Follow the prompts to finish importing.

