CODERS' CORNER

CENTRE FOR THE STUDY OF AFRICAN ECONOMIES

CREATING NETWORK DIAGRAMS IN GEPHI

Network diagrams can be helpful at the beginning of a project to understand the data, and at the end of a project to bring a result to life. Gephi is a useful program for creating, manipulating and rendering high quality graphs. It is open-source and a bit temperamental, relative to say Stata or Matlab, but more than worth the odd crash (just make sure to keep saving). It is free to download from here- https://gephi.org/users/download/.

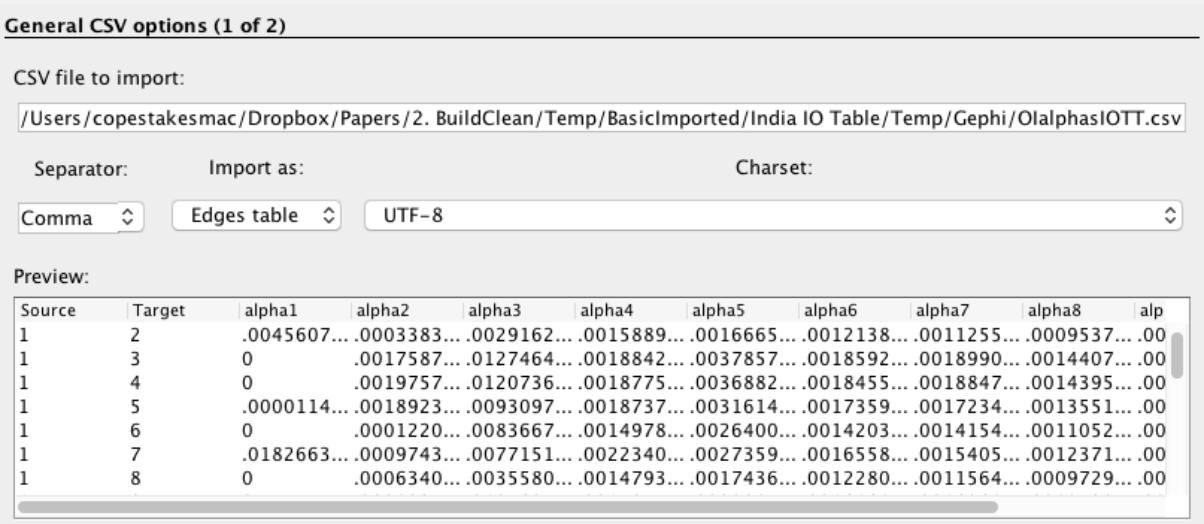
Graphs are made of nodes and edges, so first we need to import them. We need a file (.csv or similar) of edges, with the structure:

Source	Target	Weight
1	2	*x*
1	3	*y*
2	1	*z*

where Source and Target are nodes, and weight refers to the strength of the connection between them. Then we need a file of Nodes with their characteristics:

Node	Characteristic 1	Characteristic 2
1	*a*	*x*
1	*b*	*y*
2	*c*	*z*

Then start a new Gephi project and import these using the wizards, e.g.:



General CSV options (1 of 2)

CSV file to import:

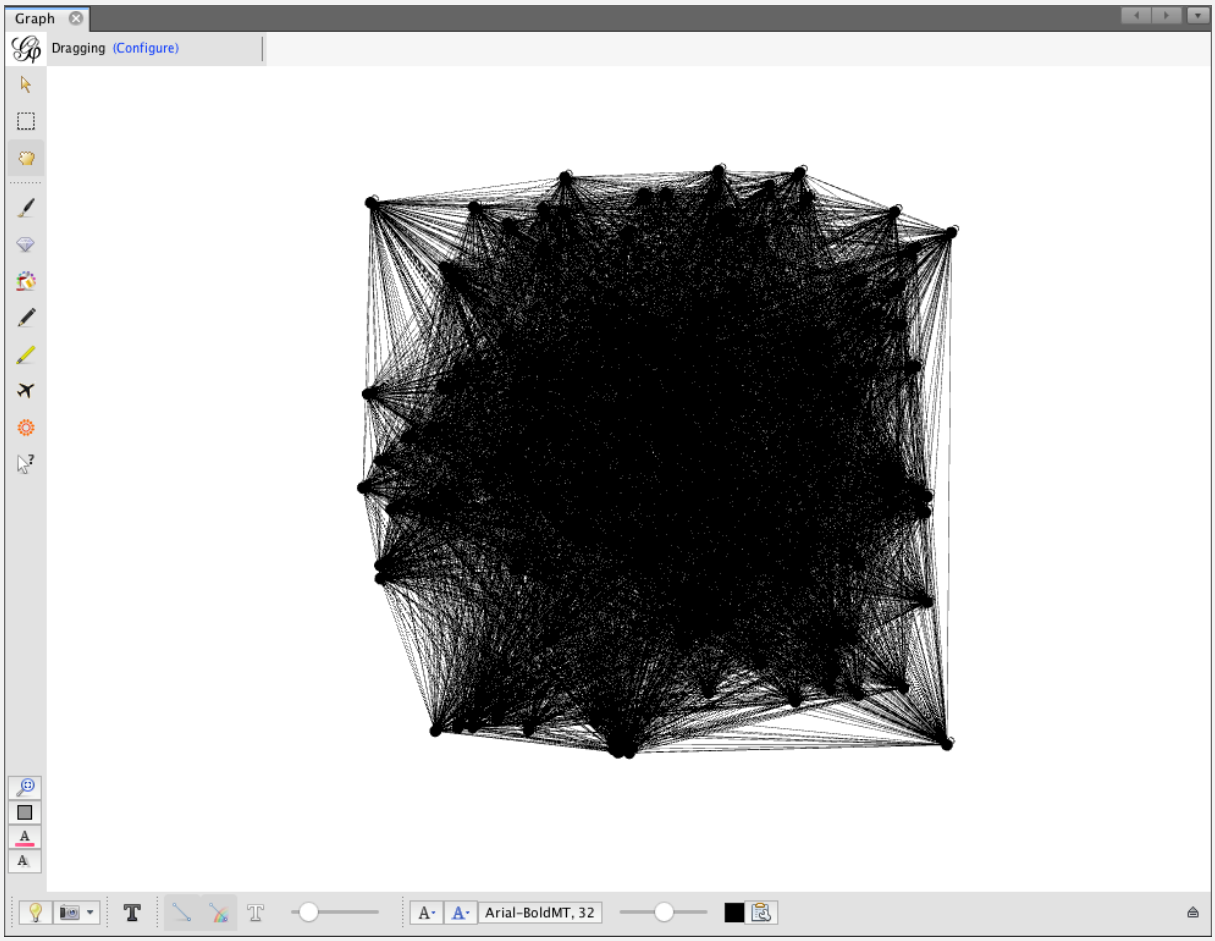
/Users/copestakesmac/Dropbox/Papers/2. BuildClean/Temp/BasicImported/India IO Table/Temp/Gephi/IImpcompsDownstreams.csv

Separator: Comma | Import as: Nodes ta... | Charset: UTF-8

Preview:

ID	Description	cTotalOutp...	OutputTari...	impcz1999	impczIV19...	Downstrea...	Downstrea...	Downstrea...	Downstrea...	Downstrea...	Dc
1	Paddy	7984225			.61724079	8.872678...	7.190924...	6.000615...	5.649291...	4.936249...	4.4
2	Wheat	5033559	0	.0000938...	.0000309...	9.848922...	7.616078...	6.649070...	6.175852...	5.426264...	4.8
3	Jowar	538810			.0108286...	9.010155...	12.72162...	9.152012...	9.245628...	7.864214...	7.1
4	Bajra	328734				9.565204...	12.36011...	9.215532...	9.274125...	7.895238...	7.1
5	Maize	549970	0	.5973435	.58913517	12.25958...	12.13931...	8.923253...	8.720153...	7.509476...	6.7
6	Gram	762143				5.754669...	9.567851...	7.273375...	7.124255...	6.098045...	5.5
7	Pulses	1379325	40	.33257619	.16496447	7.811280...	10.63919...	8.153414...	7.799538...	6.793547...	6.1

This produces an uninformative, and somewhat sinister, initial depiction:



We can now play around with all Gephi's great features to reach a representation that is useful. This example uses India's input-output table for 1998, restricted to primary commodities and manufacturing. Firstly, we only care about edges which represent a genuine supply relationship – i.e. where product A is an input to product B, or vice versa. In the Data Laboratory tab, I replace the default values of the field 'Weight' with those of the field 'alpha1', i.e. with the input shares computed from the input-output table.

Overview | Data Laboratory | Preview

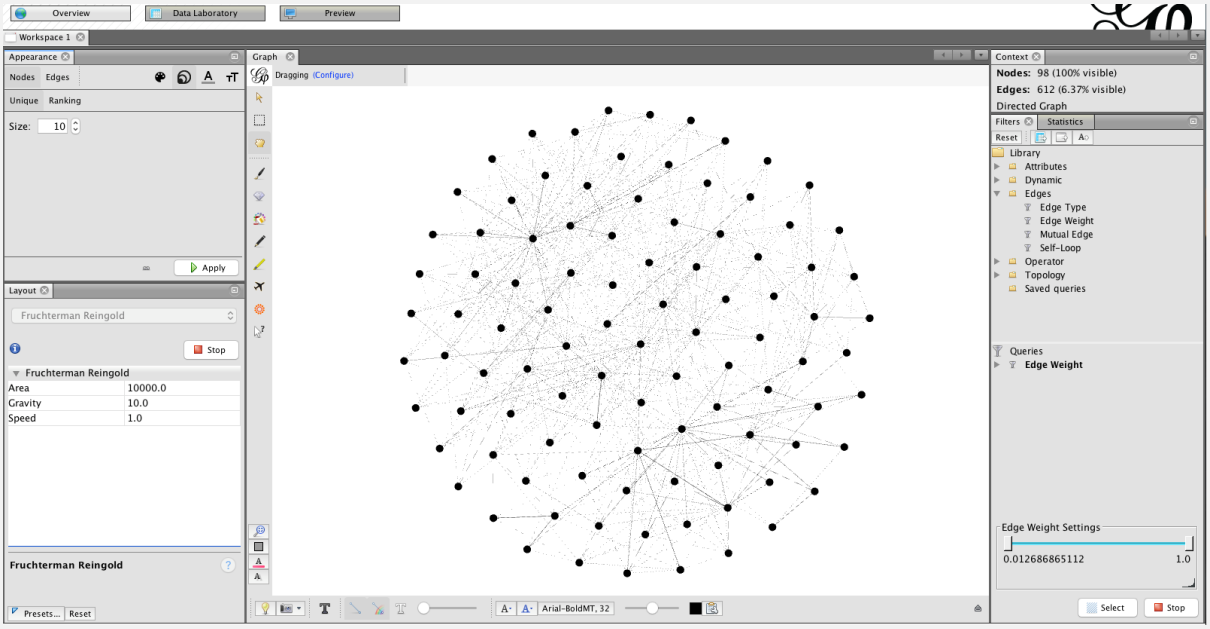
Workspace 1

Data Table

Nodes | Edges | Configuration | Add node | Add edge | Search/Replace | Import Spreadsheet | Export table

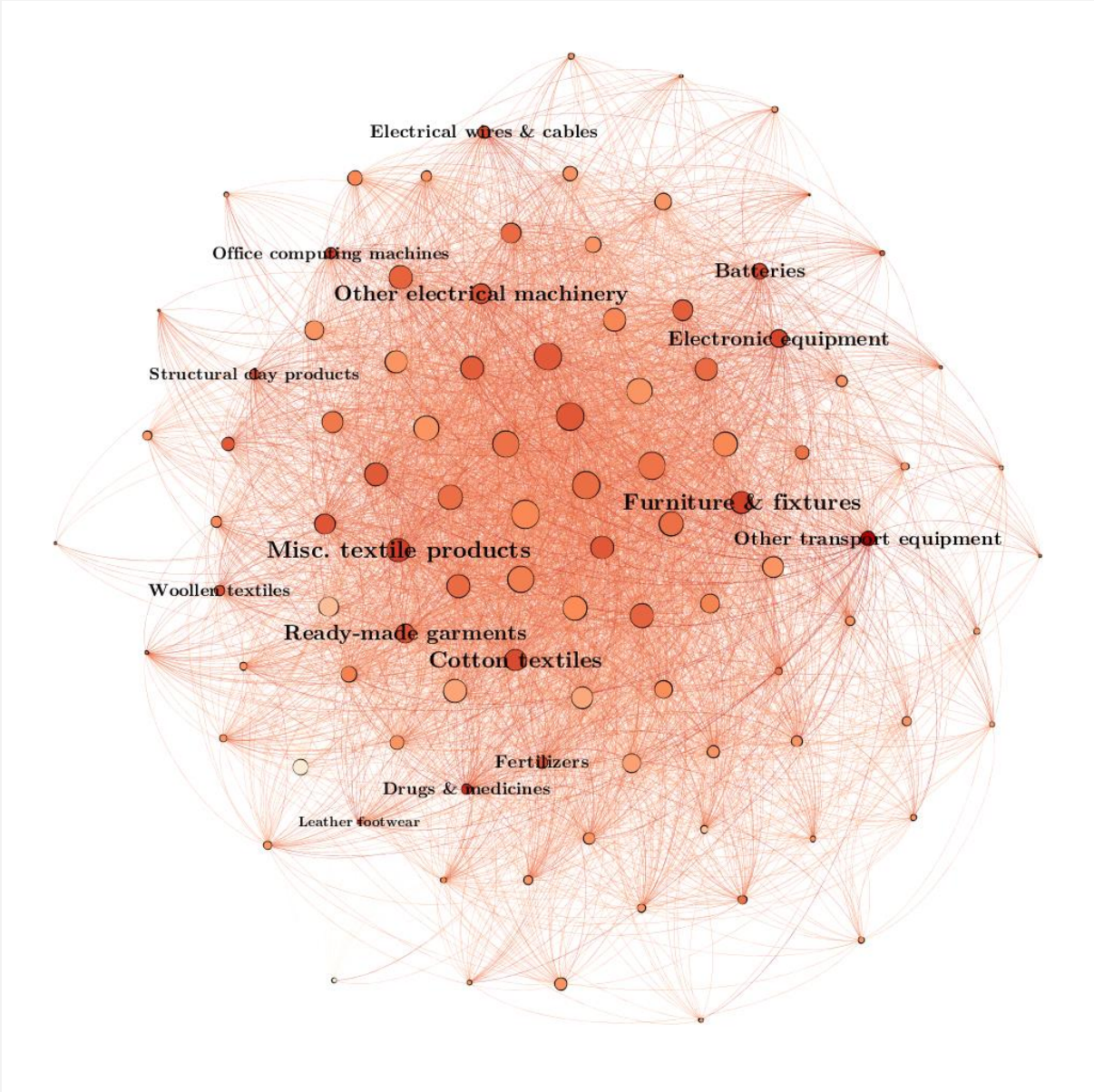
Source	Target	Type	Id	Label	Interval	Weight	alpha1
66	63	Directed	7537			0.012688	0.012688
103	61	Directed	11790			0.01273	0.01273
66	64	Directed	7538			0.012734	0.012734
58	85	Directed	6639			0.012751	0.012751
57	63	Directed	6502			0.012762	0.012762
98	53	Directed	11207			0.012762	0.012762
106	99	Directed	12173			0.012775	0.012775
24	37	Directed	2681			0.012784	0.012784
51	54	Directed	5803			0.012793	0.012793
52	97	Directed	5961			0.012842	0.012842

Returning to the Overview tab, I can then filter on Edge Weight to view only those edges with a weight greater than zero, and on the left I can run the Fruchterman-Reingold algorithm to spread the nodes out in a clearer way:



At this point, one can then play around with Gephi's various features, which are generally quite intuitive. For instance, I might size the nodes by out-degree (i.e. the number of downstream goods that they are inputs for) to get a sense of their importance in the network.

Say one is interested in understanding the impact of Chinese competition on the Indian production network. When importing the nodes table, I include the characteristic 'change between 1999 and 2013 in the share of Chinese imports in total Indian imports of the good'. I can then colour nodes darker red the greater this increase in import competition, and add labels for the 15% of products experiencing the greatest increase. Finally, switching to Gephi's Preview tab, I can export the resulting graph as a high-resolution pdf:



We can see that the main Indian industries affected by Chinese import competition were neither primary commodities (the large nodes at the centre, which supply many other industries), nor advanced manufactures (generally small and at the edge, as they are final goods). Instead the most affected industries were intermediate manufactures, with particular clusters in textiles, electronics and chemicals.

Gephi also has the "filters" feature which allows users to visualise specific kinds of nodes and edges i.e. a subgraph of the original graph. For example, the "Attributes" filter can be used to only view links among nodes that exhibit a certain characteristic, the "Edges" filter to only view edges with a certain weight , and the "Topology" filter to view the "ego network" of any pre-specified node i.e. their immediate links, the links of their links, and so on.

Alex Copesake, DPhil Candidate in Economics, Oxford
12 October 2020