

MY461: Social Network Analysis

LT 2020

Take Home Exam

Instructions to candidates

This paper contains five questions. Answer all questions. All questions will be given equal weight (20%). Responses for each question should be a maximum of 500 words excluding tables and figures. Please include a bibliography with any cited sources (this does not count towards the word count).

The exam questions will be released on April 1, 2020. The exam is due on May 13, 2020 at 16:00.

Submission will be done through Moodle. Please submit your answers in a PDF file. You will be evaluated based on your responses to the 5 prompts. To help us determine where any errors were made, however, you must additionally submit an annotated R or Rmd file that presents the code used to arrive at your responses.

Background information:

For this exam, we'll be focusing on two networks representing major trade relations between countries in 2018. The data comes from UN Data (data.un.org) and records the **top three partners** for each country, in terms of the percent of imports and exports; specifically, (1) the top three countries that each country exports goods to and (2) the top three countries that each country imports goods from, along with the percent (of imports or exports).

We are providing you with three files:

The first file (`country_metadata.csv`) has basic information on each of the countries. The variables are as follows:

- Country – the name of the country/region/area.
- ISO.alpha3.code – a three letter code for each country/region/area
- M49.code – another standard numerical coding for each country/region/area
- Region.1 – the primary region that the country is located in. (See more [here](#).)
- Region.2 – an alternative subdivision of regions (primarily joining some smaller regions into Latin America and Sub-Saharan Africa)
- Continent – the continent of the country. (Again, see more [here](#).)
- GDP.per.capita – the gross domestic product per capita in US dollars, as of 2017, again as reported by UN Data. (Note that 8 countries/regions/areas do not have a reported GDP – for these you should assign them to have the median GDP of all other countries)

The second file (`trade_imports_2018.csv`) records **the top three countries from which each country imports goods**. The first column is the focal “Ego” country and the second is the “Alter” country, from which the focal country receives imports. The third column (“PercentOfImports”) records the percent of the focal country’s imports come from that country (so, e.g., 19.4% of Afghanistan’s imports come from Iran, making it their top import partner). In a few cases, countries do not have a full set of three partners, as it was somehow unspecified.

The third file (`trade_exports_2018.csv`) records **the top three countries to which each country exports goods**. It is structured in a similar way to the previous file, with the focal “Ego” country, the “Alter” country to which they export goods, and the “PercentOfExports” that are sent to that particular country. (So, e.g., 47.5% of Afghanistan’s exports go to Pakistan, making it their top export partner).

Using these files, create two trade networks, one representing import partnerships and one representing export partnerships. Make the percentage (whether of imports or of exports) an edge attribute and the country metadata as vertex attributes.

With these networks in hand, answer the following questions.

- 1 Consider the overall metrics (density, average path length, transitivity, and reciprocity) of the import and export networks. How do the two networks compare to each other? Compare the import and export networks to random networks created with the Erdős–Rényi and the Barabási-Albert models. What do these comparisons tell you about the nature and structure of the relationships among the countries? In your answer, make sure to define each of the metrics and give an intuitive interpretation for them.
- 2 Which do you see as the most influential countries in the trade networks? Identify two potential meanings of “influence,” as proxied by different centrality measures. Justify your choice of each centrality measure. In that justification, present clear interpretations of what each centrality measure is capturing about the position of the countries. Calculate both of your chosen centrality measures on the import network and on the export network and identify the country that has the highest value for each. Discuss the patterns you see when you compare between the import and export networks and what it implies about the identified countries and about trade relations, generally. Make explicit reference in your response to concepts related to node centrality covered in the course material. [Note that if you use edge weight for a centrality measure, you need to identify how the calculation interprets those weights (i.e., do higher values mean greater closeness or greater distance?); in those cases where the measure assumes that higher values means greater distance, you should use $1/\text{weight}$ in the calculation.]
- 3 How does geography influence trade relations? Calculate the assortativity by continent and the probability of an import and export tie within and between each continent (using a blockmodel approach) for the two trade networks. Run the spinglass community detection algorithm on the two trade networks (do not use edge weight). Plot both the import and export networks twice with nodes coloured by (1) the results of the community detection algorithm and (2) continent. Make sure that your plots are legible and informative; include a figure label, size the nodes by one of the centrality measures you chose in response to the second prompt, and if you choose to label the nodes, make sure the network isn’t obscured. Discuss how the resulting communities align (or not) with continent and what the assortativity, blockmodel, and community detection results imply about trade relations.
- 4 Do different countries seem to fulfil different roles in the trade networks? Evaluate the structural equivalency of the countries using both the import and export networks. Do so using the `equiv.clust()` function in the `sna` package (with the code `equiv.clust(list(imp_net, exp_net))`, where `imp_net` is your import network object and `exp_net` is your export network object). Plot the dendrogram (making sure that the countries are identified in the plot) and use the results of the structural equivalency calculation to divide the countries into six equivalency classes. Interpret the six groups, drawing on prior analyses. Discuss what this implies about the countries, drawing on the concepts of position and role.
- 5 What helps predict whether one country exports goods to another? Run the following exponential random graph model for the export network:

```
model <- ergm(exp_net ~ edges + nodeifactor("continent") + nodematch("continent") + nodeicov("log_gdp") + mutual)
```

(Note that “log_gdp” should be the log of the per capital GDP for each country). Interpret each term in the ERGM (except edges) – how does each term influence whether or not a country is connected

to another (use odds ratios in your substantive interpretation of each term)? In your discussion, make explicit reference to the concepts of homophily and reciprocity. Propose one additional term to add to the model and provide a justification for it. Make that addition and evaluate the effect of that term and how the model changes with its addition.