Four empirical networks

Junhui Cai, Dan Yang, Wu Zhu, Haipeng Shen, Linda Zhao

In this report, we show that many (economic) networks are, in fact, low-rank and each of their leading singular value dominates the non-leading ones. Specifically, we examine the spectral properties of four empirical networks: (A) global trade network, (B) innovation network, (C) production network, and (D) equity network. We will first describe each network and show the top 20 singular values of each network in 2013.

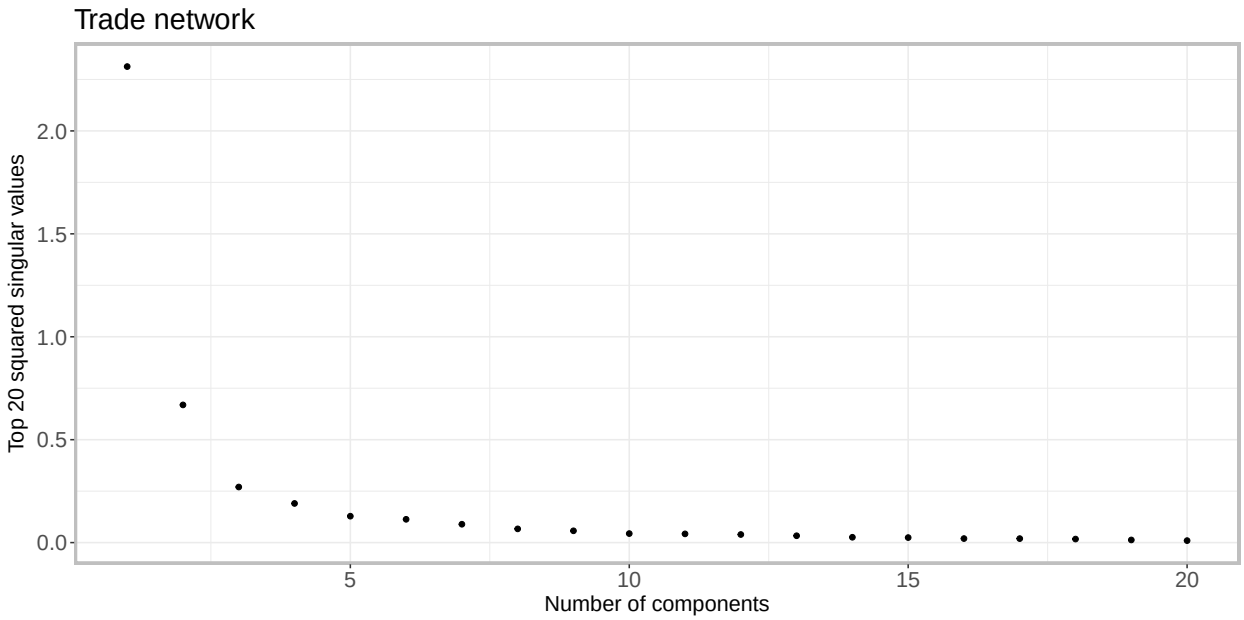
To reproduce this report, `Knit` this file in RStudio. One can set `echo = T` for the each chunk or globally `knitr::opts_chunk$set(echo = T)` to show the code in the report.

1 Data

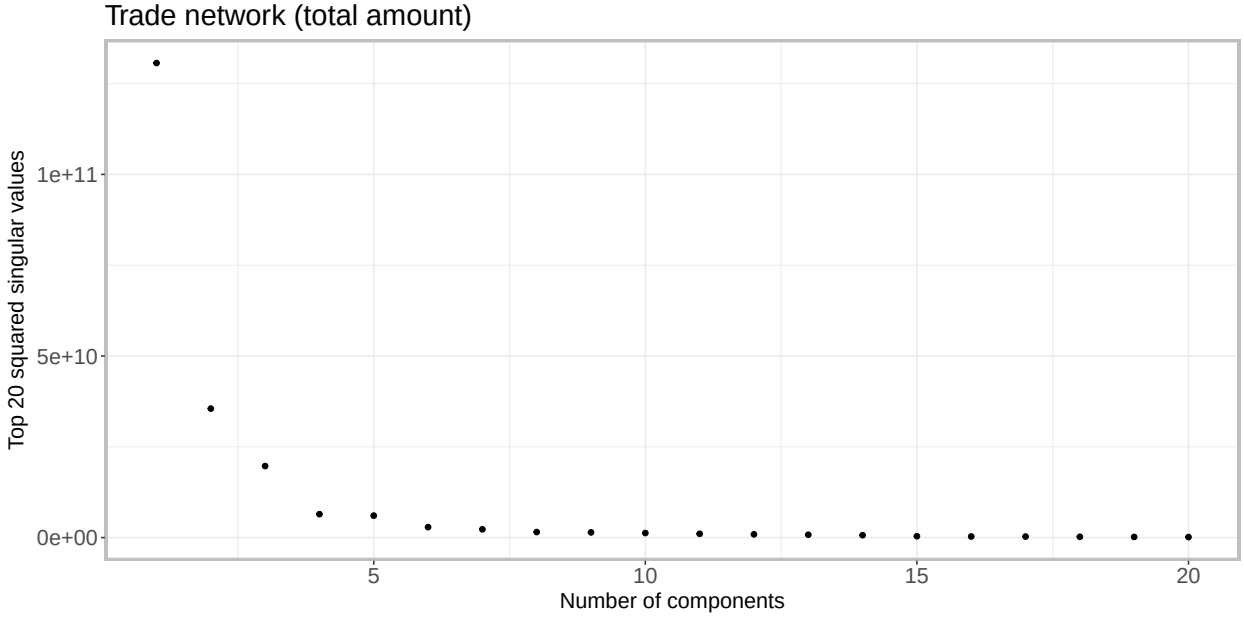
- `data_empirical_network/trade_data_sub.csv`: Trade network
- `data_empirical_network/naics3_uspc_naics3_W_matrix.csv`: Innovation network
- `data_empirical_network/IO_naics3_1997_2018_naics07.xlsx`: Production network
- `data_empirical_network/equity_network_svd.csv`: Equity network is proprietary data but we provide the top 50 singular values

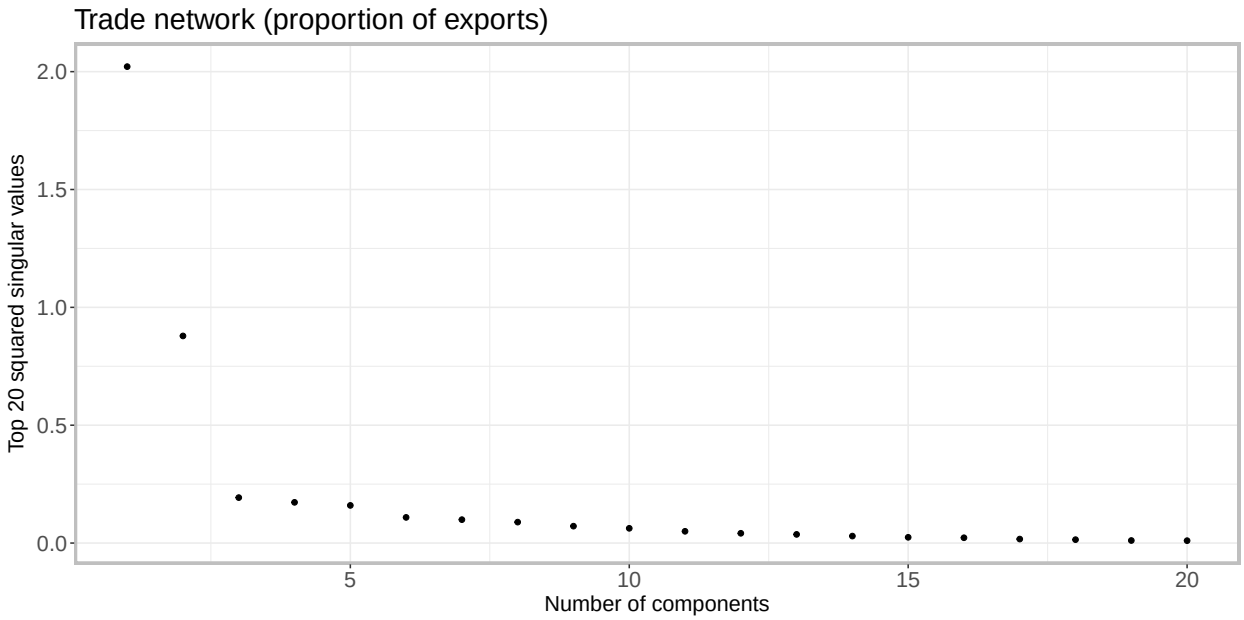
1.1 Trade network

The global trade network is the country-level trade network. We include all 35 countries here in the bilateral trade data from the correlates of war project (COW). The trade network is defined as the proportion of imports from other countries, i.e., the proportion of imports from country i among all the imports of country j.



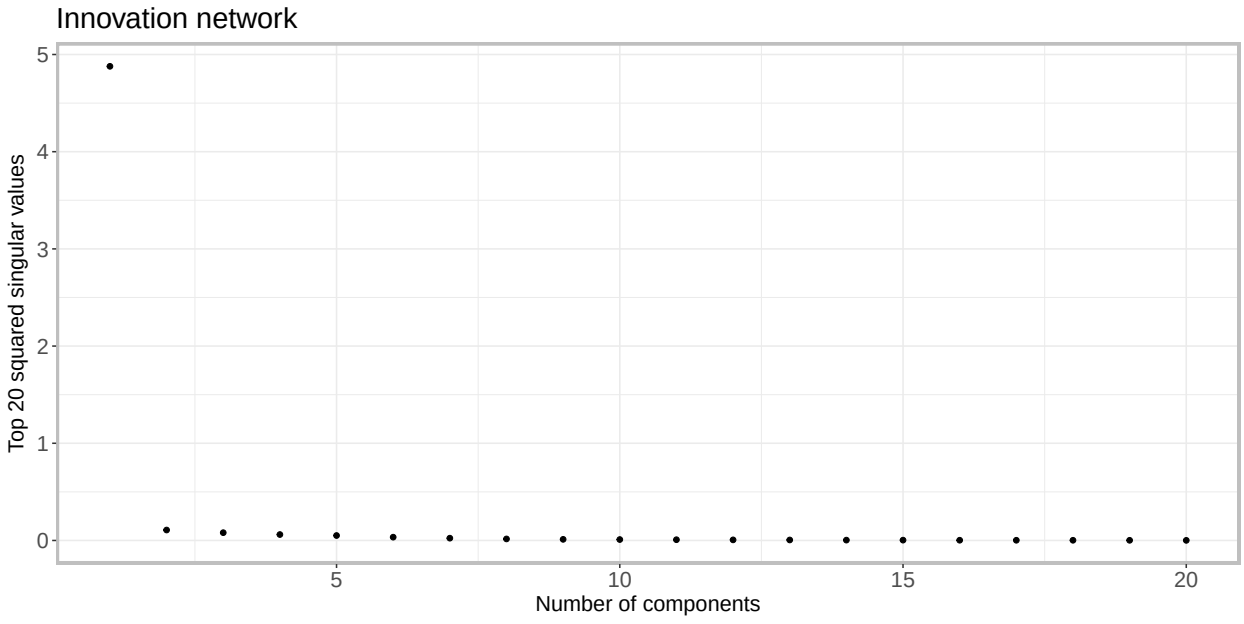
Similarly we define the trade network as the import-export amount in US dollars or as the proportion of exports to other countries. The results are similar.





1.2 Innovation

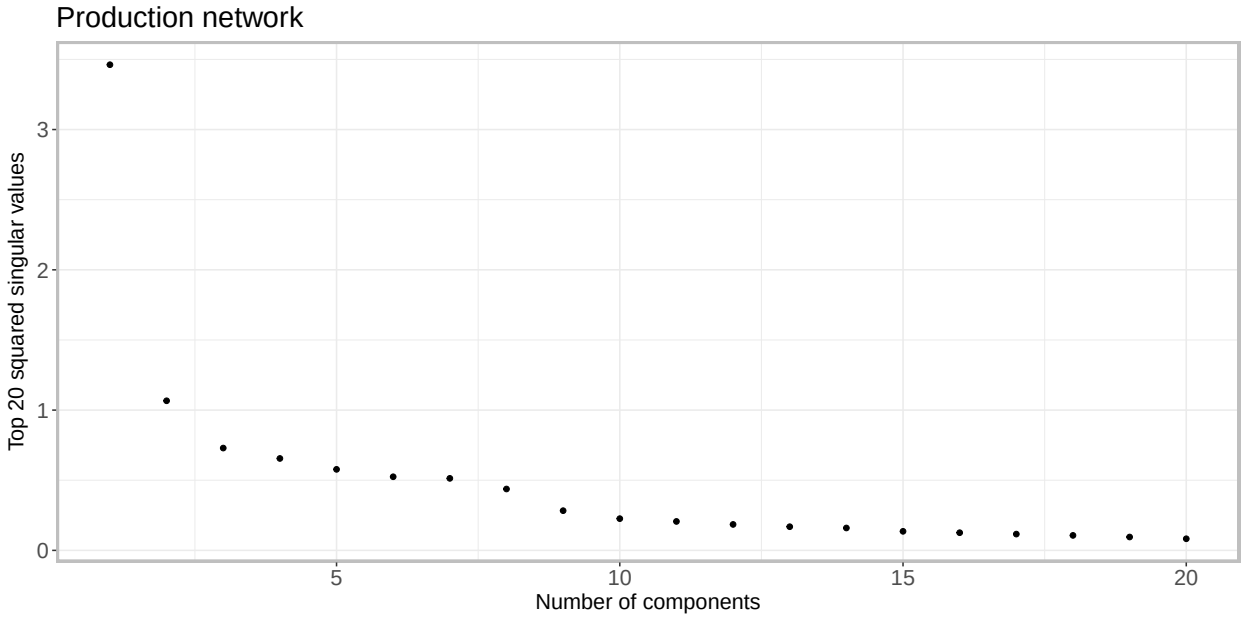
The innovation network is an industry-level network of knowledge flow based on patent citations in the US. The industry is classified by the 3-digit code of the North American Industry Classification System (NAICS) and we cover 87 industries in 2013. Please refer to the Data Appendix of Zhu and Yang (2020) for a detailed description of the construction of the innovation network.



1.3 Production network

The production network is a network of the input-output flow at the industry level in the US. Particularly, the input-output flow from industry i to j is defined as the proportion of input from industry i among all

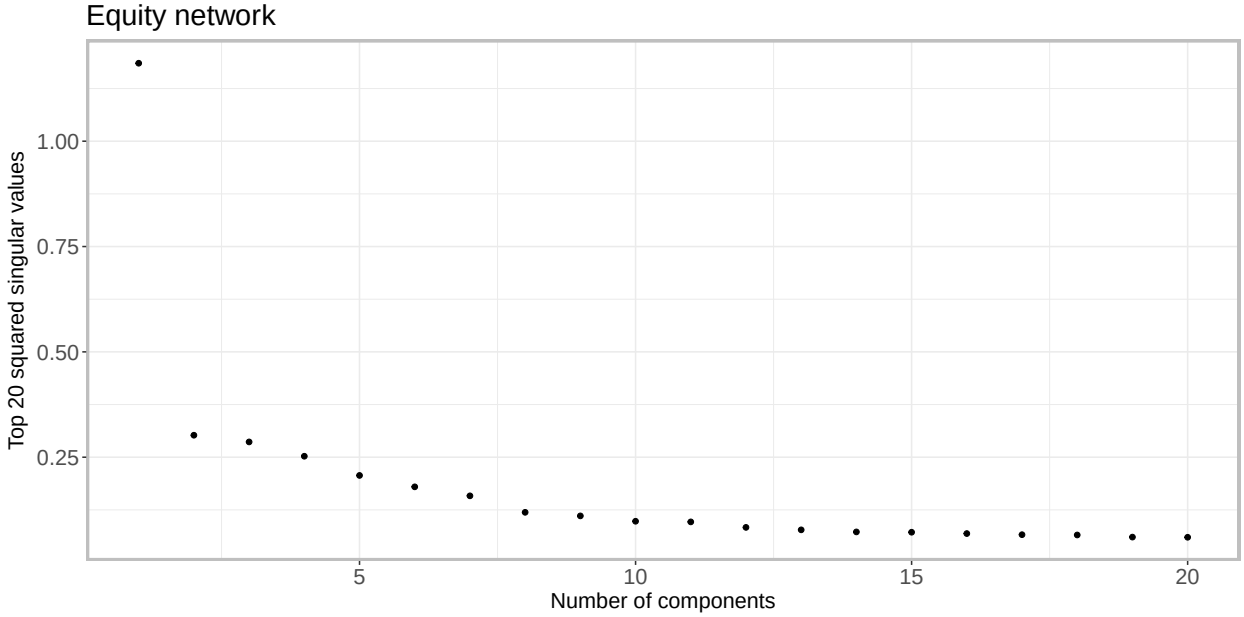
the inputs of industry j. Same as the innovation network, we use the 3-digit NAICS code and cover 87 industries in 2013. Please refer to the Data Appendix of Zhu and Yang (2020) for a detailed description of the construction of the production network.



1.4 Equity network

The equity network is a network of firm-level investor-investee shareholding relationships in China, i.e., the percentage of shares of firm j owned by firm i. In 2013, we include 3.6 million firms that have at least one firm as their shareholders2 and we focus on the largest connected component3 which includes 1.3 million firms. Please refer to Cai et al. (2021) for a detailed description of the construction of the equity network.

The equity network data is proprietary but we provide the top 50 singular values.



2 All together

The following chunk reproduces Figure S2.