

Appendix

Supplementary Document for “Estimation of Regional Economic Development Indicator from Transportation Network Analytics”, *Scientific Reports*. <https://doi.org/10.1038/s41598-020-59505-2>

Bin Li¹, Song Gao^{2*}, Yunlei Liang², Yuhao Kang², Timothy Prestby², Yuqi Gao², Runmou Xiao¹

1. School of Automobile, Chang'an University, Xi'an, Shaanxi, 710064, China

2. Geospatial Data Science Lab, Department of Geography, University of Wisconsin-Madison, WI, 53706, USA

*Corresponding author email: song.gao@wisc.edu

Modeling the GDP distribution

The empirical histogram distribution of the economic development indicator (GDP) values across all cities in the study period is a positively skewed distribution with a long-tail end (Fig. S1a). By fitting the data observations into four theoretical models (i.e., normal, gamma, log-normal, and Weibull distributions) using the maximum likelihood estimation⁷¹, the results show that the log-normal model is the best with the maximum log-likelihood of -994.39 and the minimum Akaike information criterion (AIC) value of 1992.78. AIC is used to determine which model performs the best regarding the goodness of fit and the simplicity of a model using information theory⁷². Meanwhile, the log-likelihood and AIC values for other models are gamma distribution (log-likelihood=-1002.82 & AIC=2009.64), Weibull distribution (log-likelihood=-1004.23 & AIC=2012.45), and normal distribution (log-likelihood=-1066.17 & AIC=2136.35) respectively. Fig. S1b and Fig. S1d show the curves of empirical and theoretical cumulative distribution functions of these models and their quantile-quantile plots. The skewness-kurtosis plot (i.e., the Cullen and Frey graph⁷³) is used to examine the characteristics of the skewness and the degree of tailedness (i.e., kurtosis) of the GDP data with uncertainty compared with the theoretical models. As shown in the S1c, the bootstrapping values fit better with the log-normal, gamma and Weibull models rather than the normal, uniform and logistic models. The consistent positive skewness and kurtosis values verify the "heavy-tailed on the right" characteristic of the city GDP distribution in the study areas.

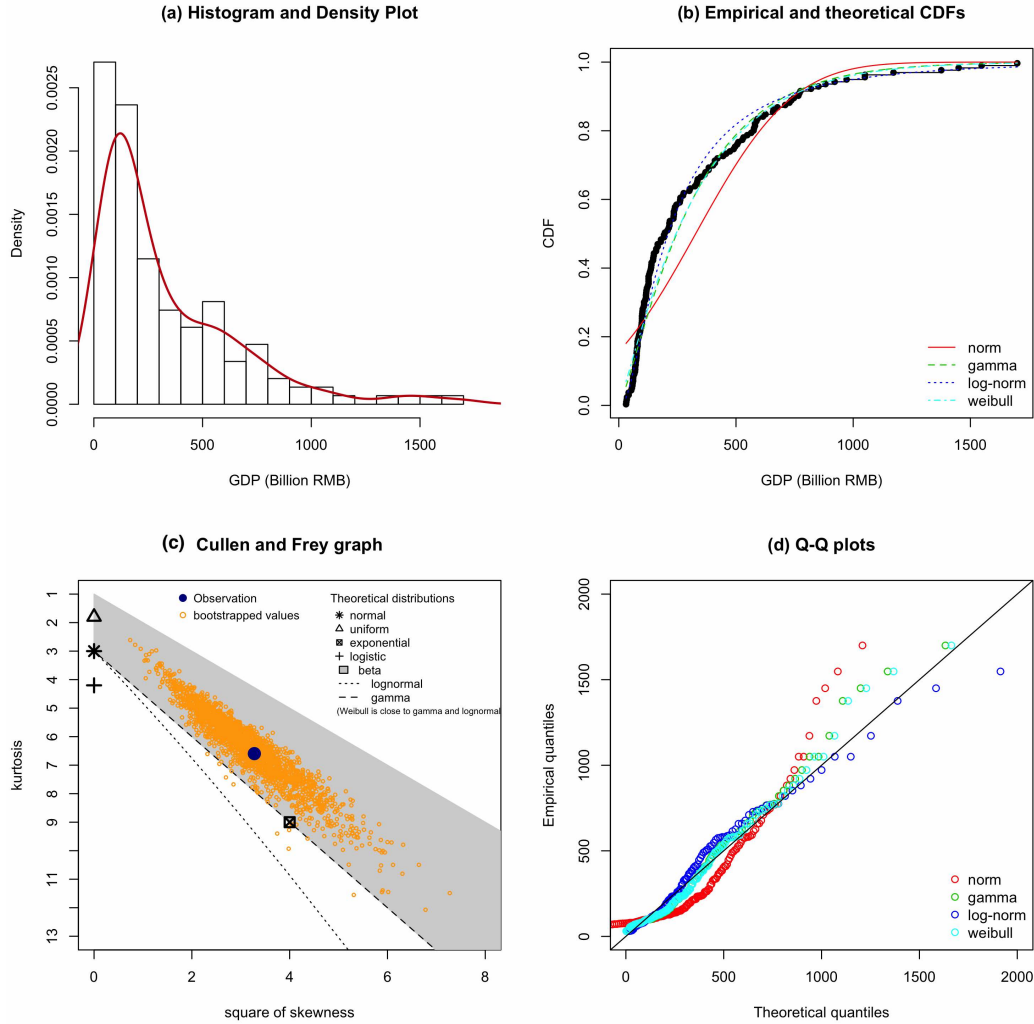


Figure S1. The empirical and theoretical distributions of all city GDP values. (a) the histogram of city GDPs; (b) the observed empirical cumulative distribution function (CDF) in black and four theoretical CDFs; (c) the Cullen and Frey graph shows the probability distribution fittings; (d) the Q-Q plot compares the empirical and theoretical quantile distributions.

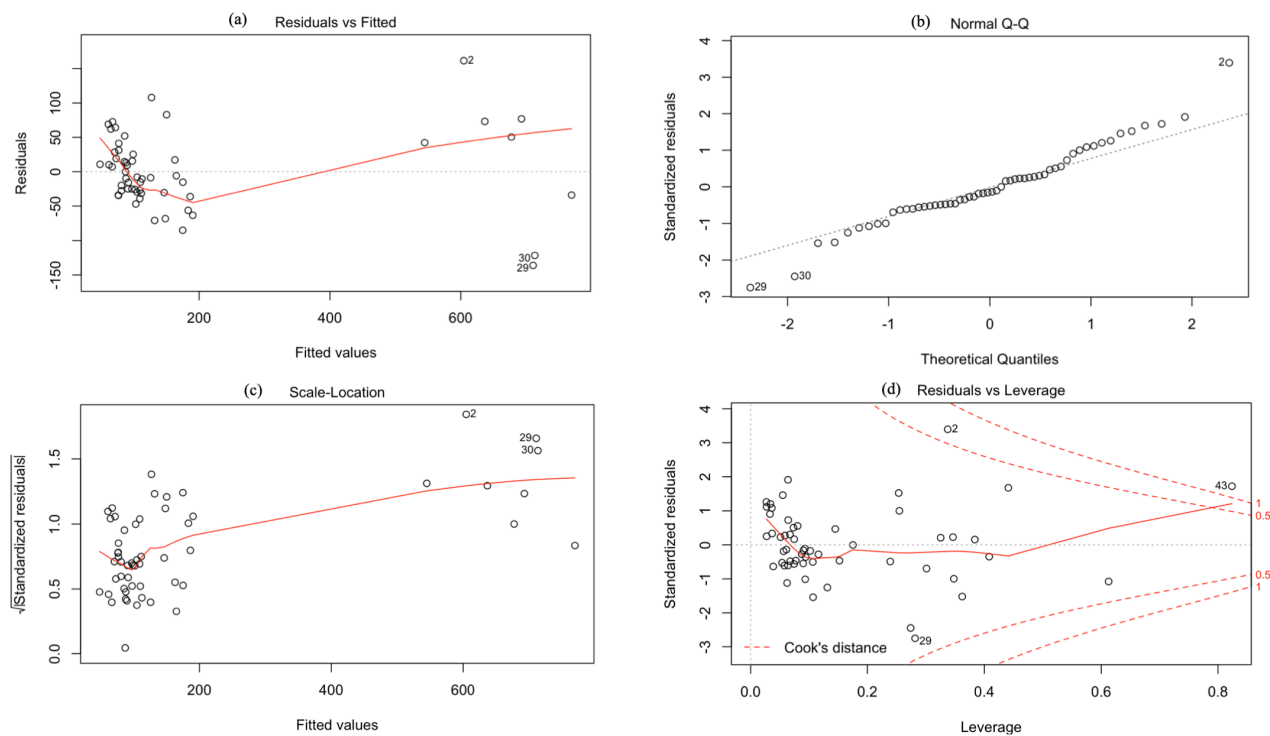


Figure S2. The residual plots of predicting city GDP values in Liaoning province using the multiple linear regression. (a) the scatter plot of residuals and fitted values; (b) the Normal Quantile-Quantile plot for standardized residuals; (c) the Scale-Location plot to check the residual spread; (d) the Residuals vs Leverage plot to find influential samples if any.

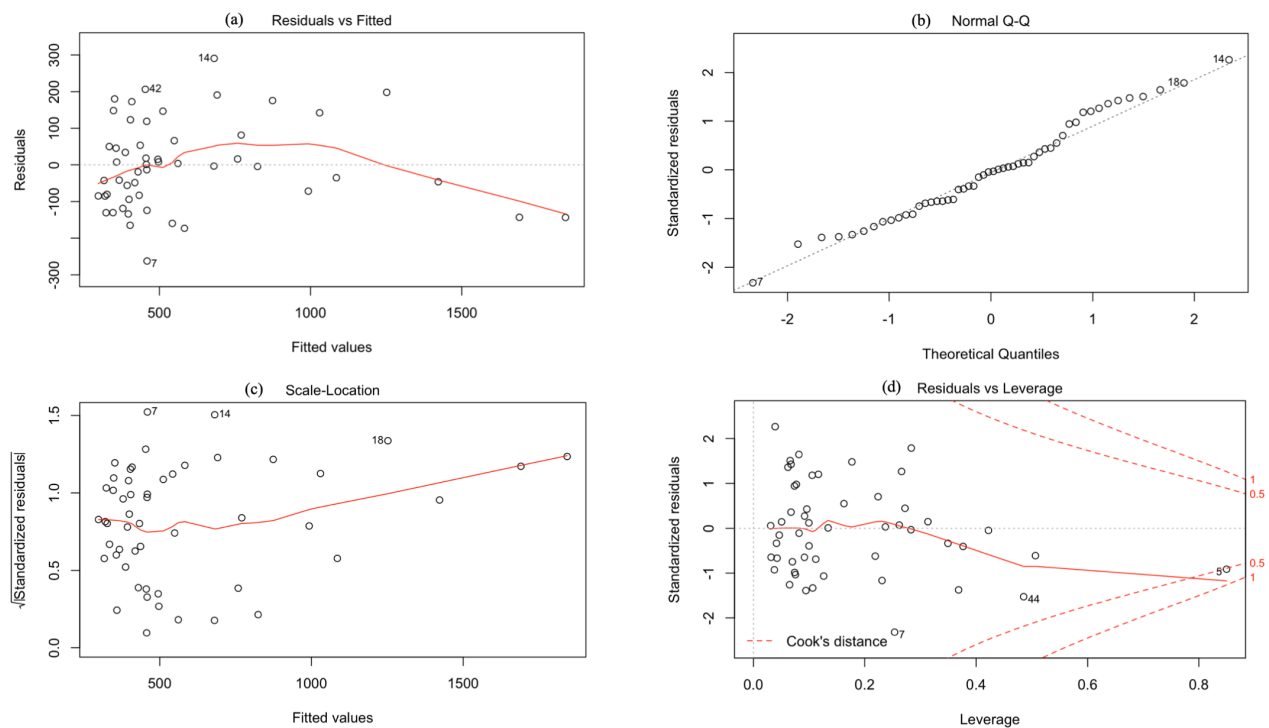


Figure S3. The residual plots of predicting city GDP values in Jiangsu province using the multiple linear regression. (a) the scatter plot of residuals and fitted values; (b) the Normal Quantile-Quantile plot for standardized residuals; (c) the Scale-Location plot to check the residual spread; (d) the Residuals vs Leverage plot to find influential samples if any.

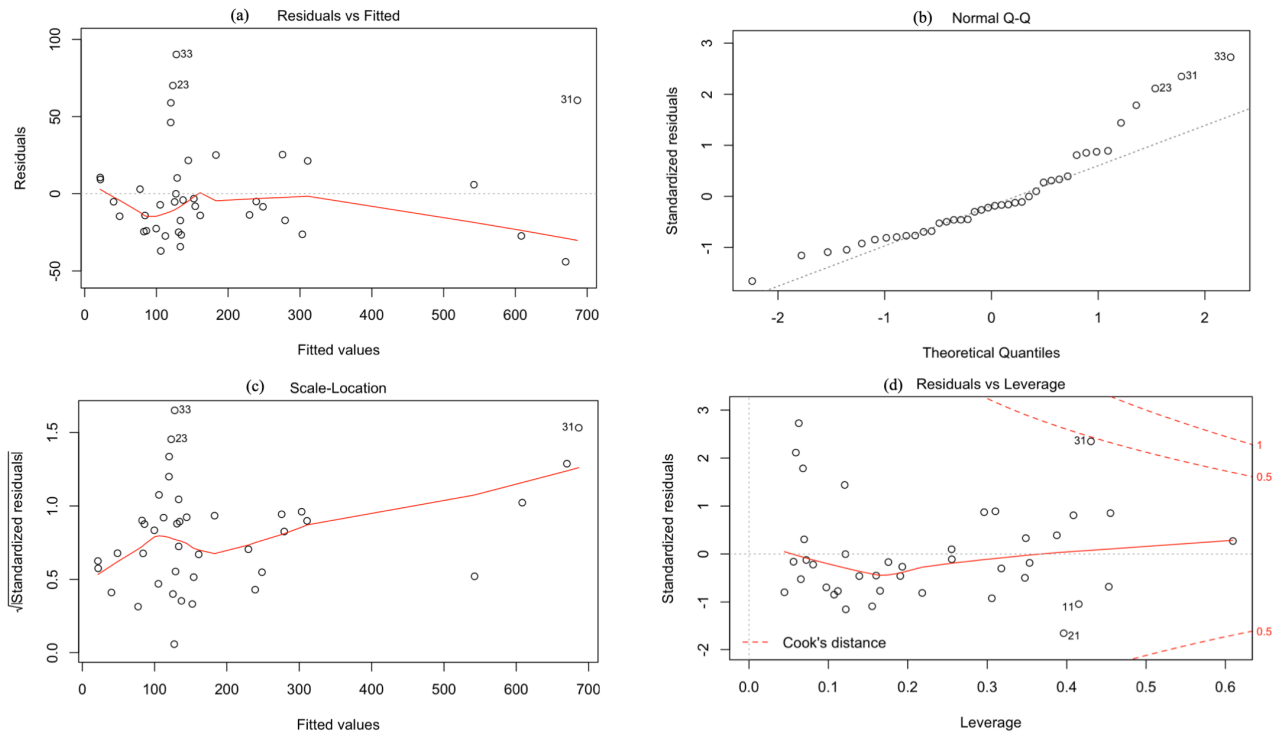


Figure S4. The residual plots of predicting city GDP values in Shaanxi province using the multiple linear regression. (a) the scatter plot of residuals and fitted values; (b) the Normal Quantile-Quantile plot for standardized residuals; (c) the Scale-Location plot to check the residual spread; (d) the Residuals vs Leverage plot to find influential samples if any.

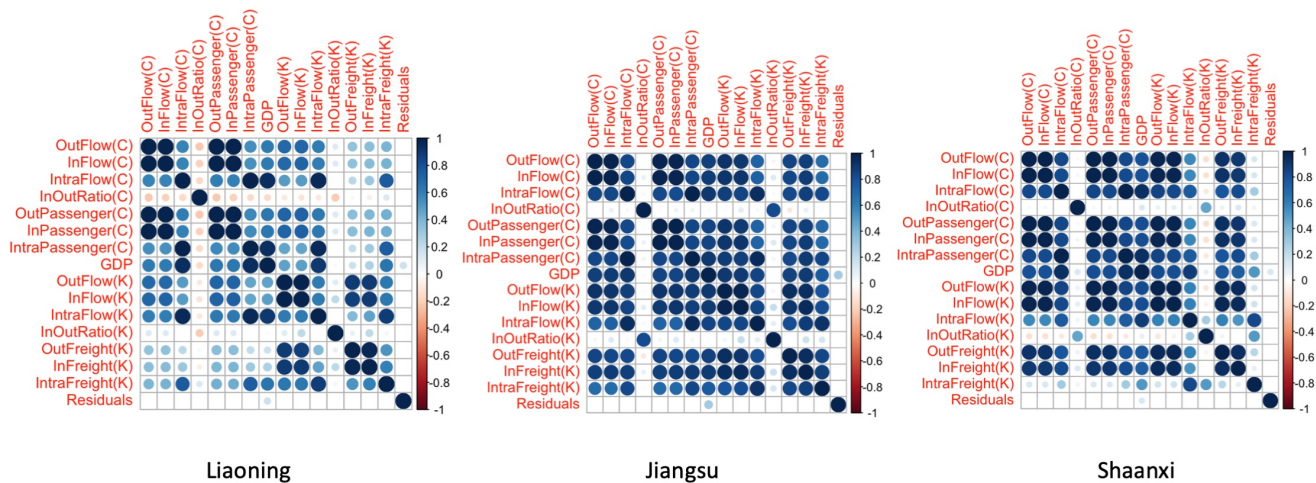


Figure S5. The visualization of correlation matrix of predictors and residuals in each province.

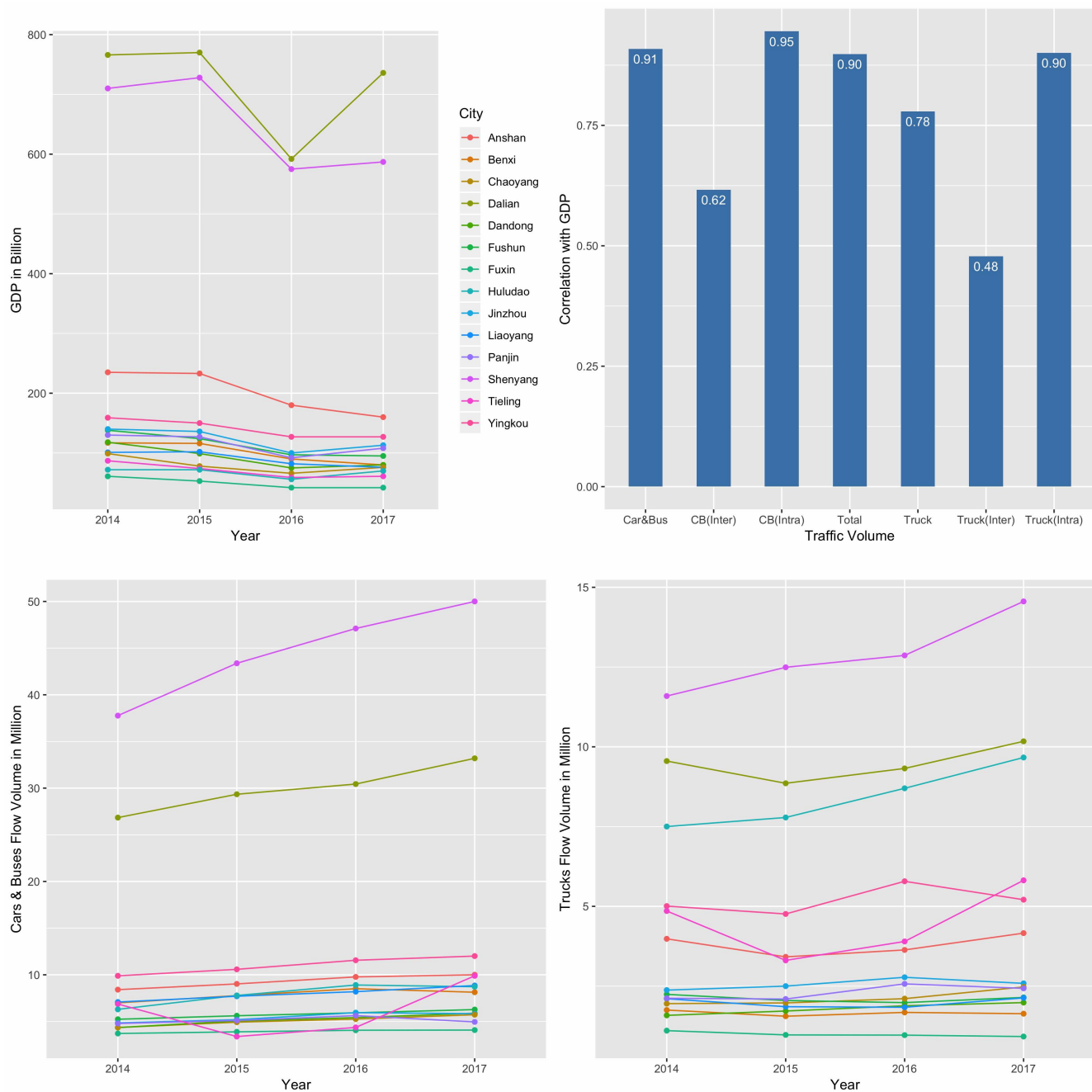


Figure S6. (a) The temporal changes of cities' GDP values; (b) the correlation between city GDP and traffic volumes; (c) the temporal changes of traffic volumes of cars and buses; (d) the temporal changes of traffic volumes of trucks in Liaoning province.

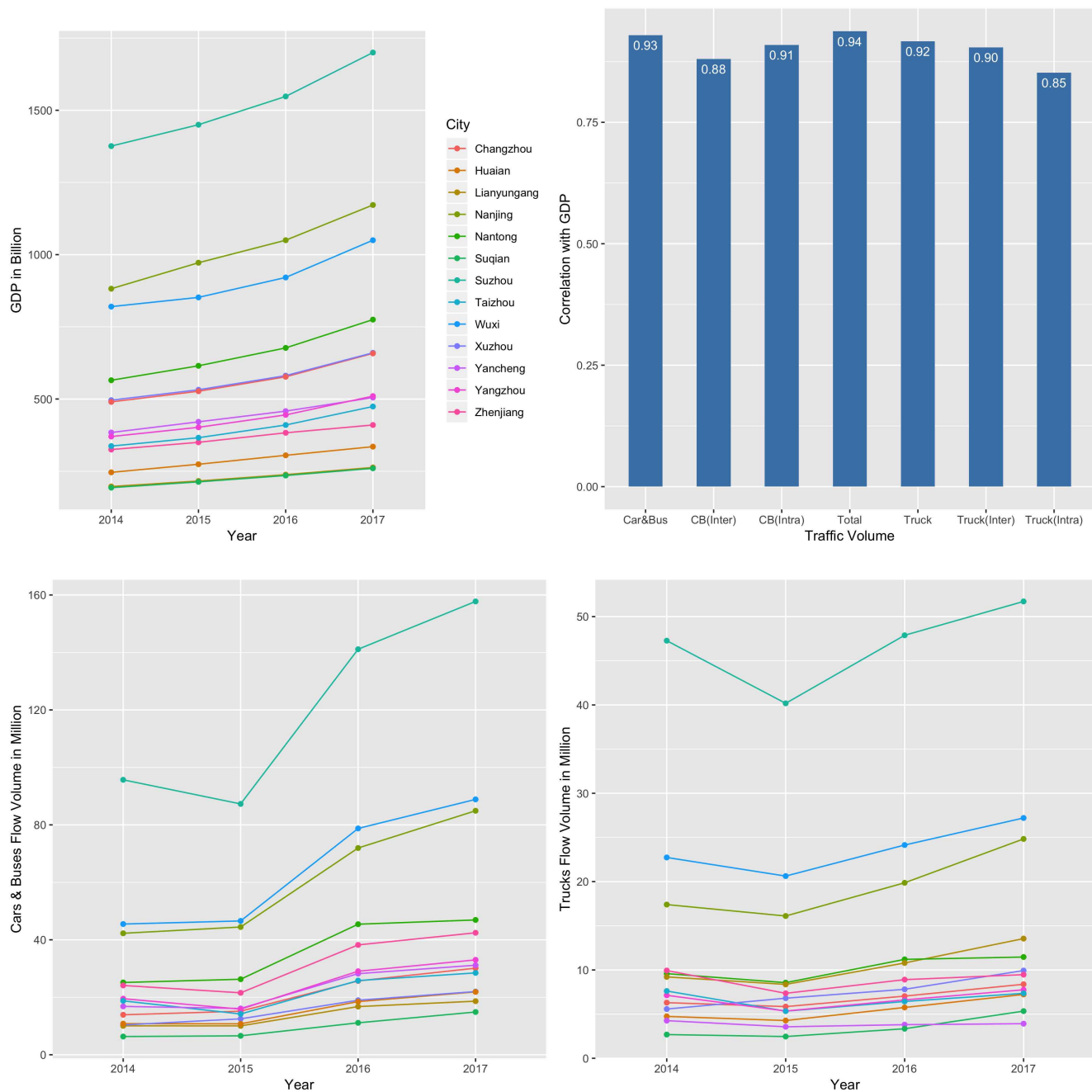


Figure S7. (a) The temporal changes of cities' GDP values; (b) the correlation between city GDP and traffic volumes; (c) the temporal changes of traffic volumes of cars and buses; (d) the temporal changes of traffic volumes of trucks in Jiangsu province.

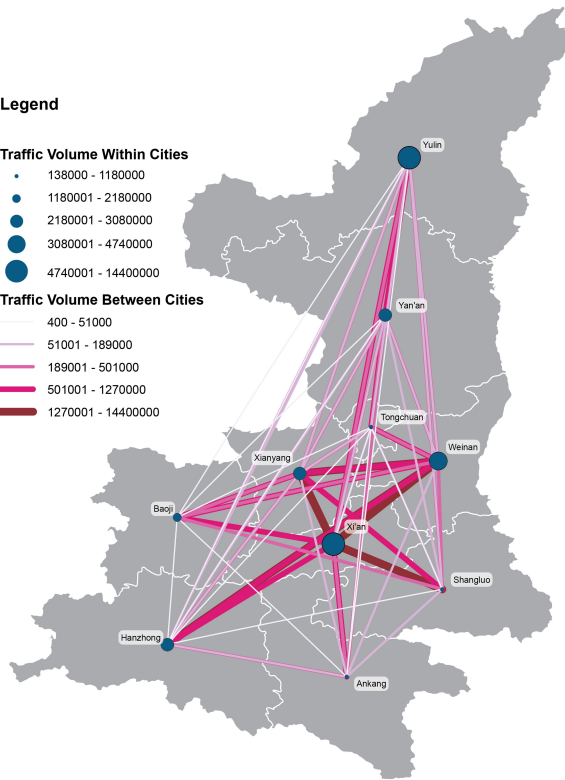
Legend

Traffic Volume Within Cities

- 138000 - 1180000
- 1180001 - 2180000
- 2180001 - 3080000
- 3080001 - 4740000
- 4740001 - 14400000

Traffic Volume Between Cities

- 400 - 51000
- 51001 - 189000
- 189001 - 501000
- 501001 - 1270000
- 1270001 - 14400000



(a) 2014

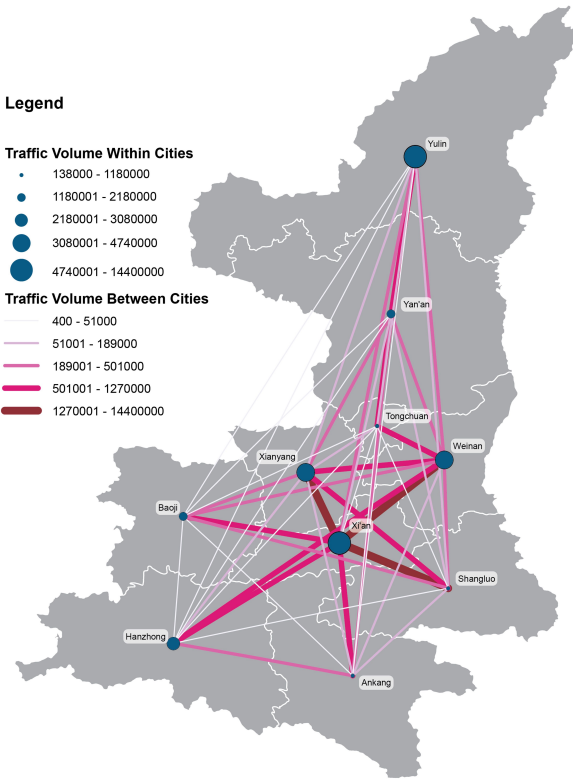
Legend

Traffic Volume Within Cities

- 138000 - 1180000
- 1180001 - 2180000
- 2180001 - 3080000
- 3080001 - 4740000
- 4740001 - 14400000

Traffic Volume Between Cities

- 400 - 51000
- 51001 - 189000
- 189001 - 501000
- 501001 - 1270000
- 1270001 - 14400000



(b) 2015

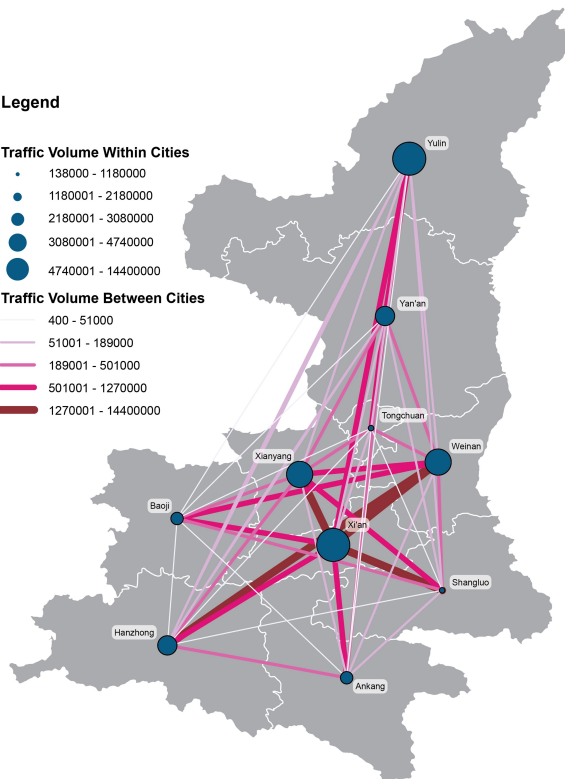
Legend

Traffic Volume Within Cities

- 138000 - 1180000
- 1180001 - 2180000
- 2180001 - 3080000
- 3080001 - 4740000
- 4740001 - 14400000

Traffic Volume Between Cities

- 400 - 51000
- 51001 - 189000
- 189001 - 501000
- 501001 - 1270000
- 1270001 - 14400000



(c) 2016

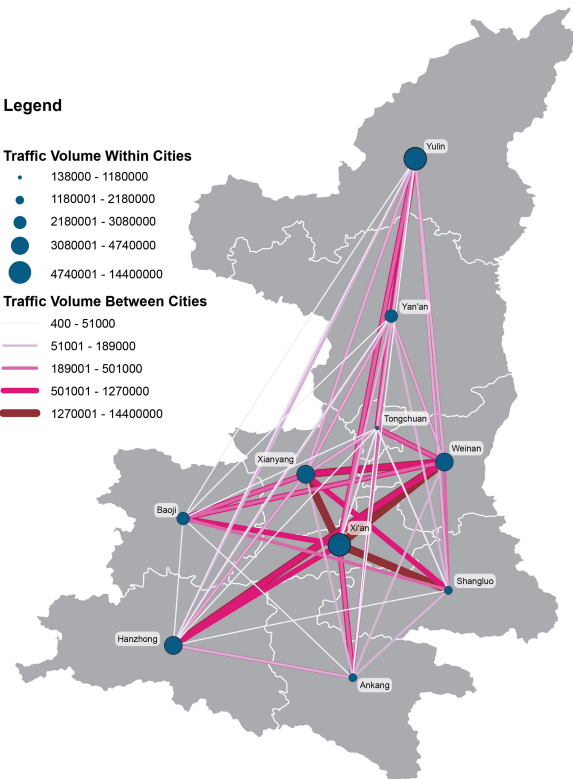
Legend

Traffic Volume Within Cities

- 138000 - 1180000
- 1180001 - 2180000
- 2180001 - 3080000
- 3080001 - 4740000
- 4740001 - 14400000

Traffic Volume Between Cities

- 400 - 51000
- 51001 - 189000
- 189001 - 501000
- 501001 - 1270000
- 1270001 - 14400000



(d) 2017

Figure S8. Traffic Volume of Freight Trucks in Shaanxi. Note: The maps were generated using ArcMap version 10.6 and Adobe Illustrator CC version 20.

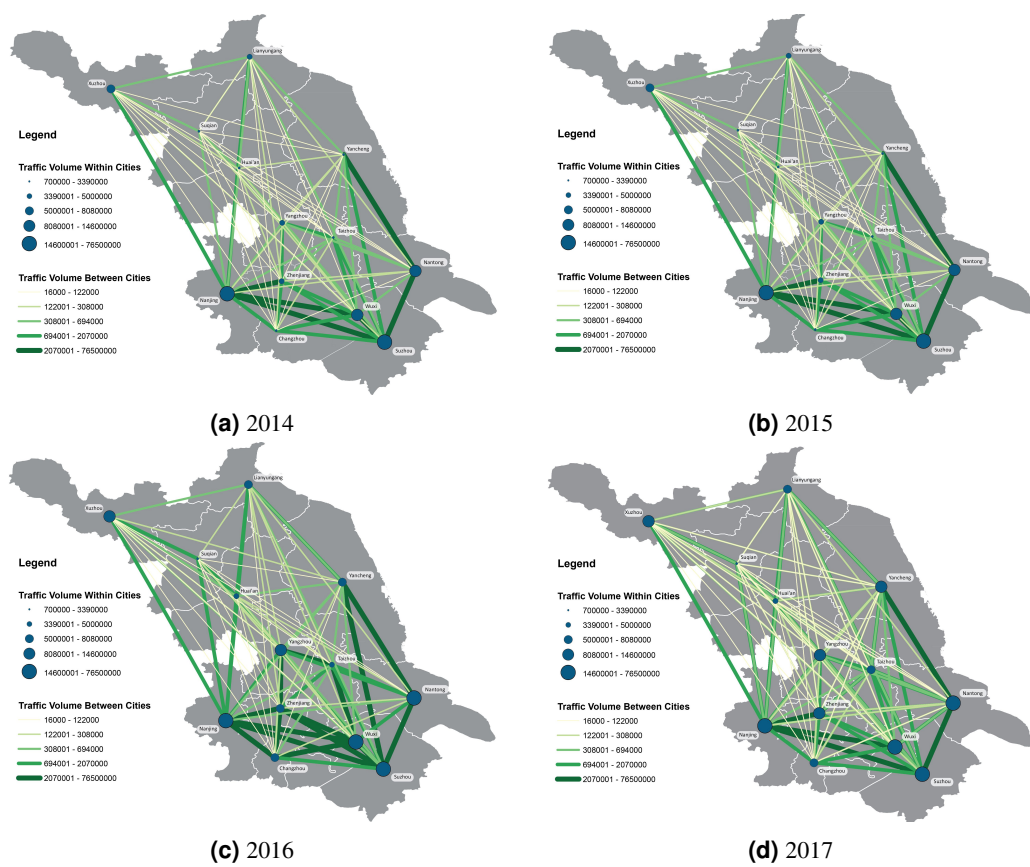


Figure S9. Traffic Volume of Cars and Buses in Jiangsu. Note: The maps were generated using ArcMap version 10.6 and Adobe Illustrator CC version 20.

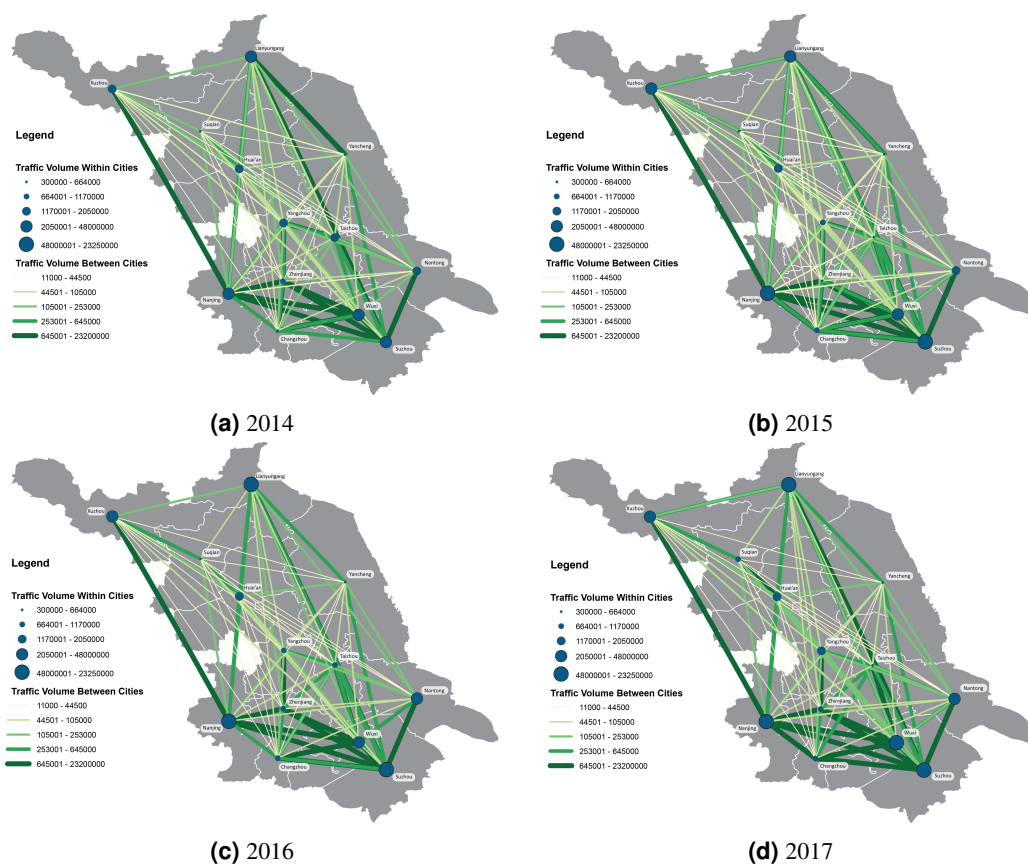


Figure S10. Traffic Volume of Freight Trucks in Jiangsu. Note: The maps were generated using ArcMap version 10.6 and Adobe Illustrator CC version 20.

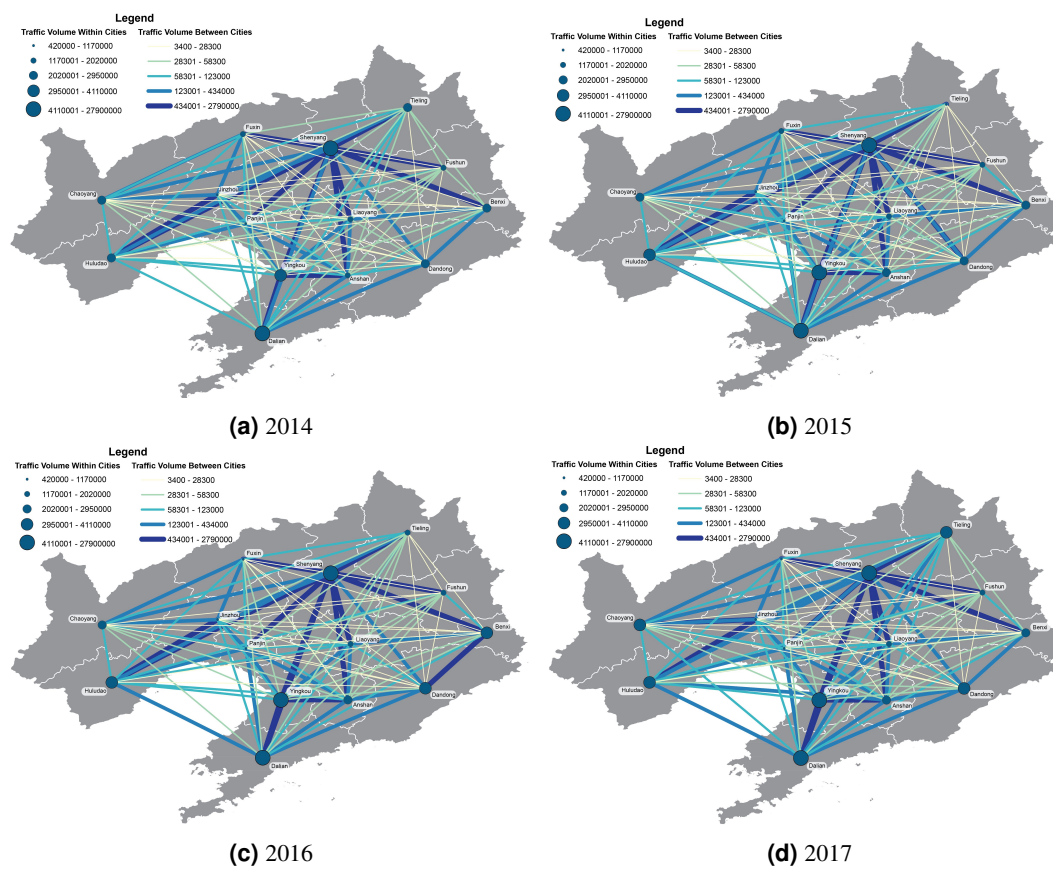
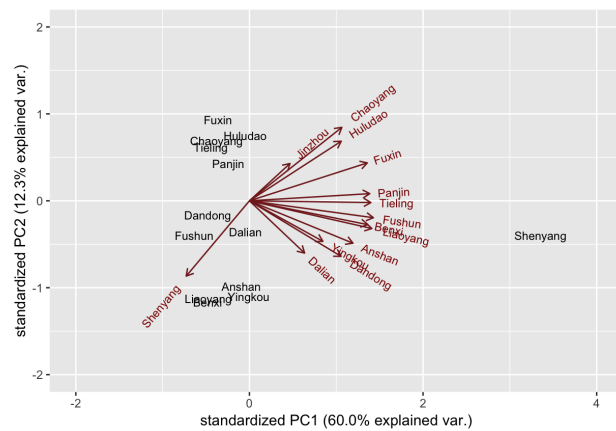


Figure S11. Traffic Volume of Cars and Buses in Liaoning. Note: The maps were generated using ArcMap version 10.6 and Adobe Illustrator CC version 20.



First PC with loading >0.3 and score > 1

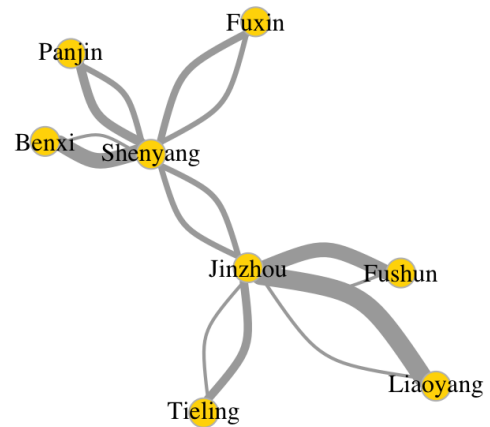


Figure S14. The PCA analysis results of spatial interaction networks of buses and cars in Liaoning. Note: The figures were generated using RStudio version 1.2.

Supplementary References:

71. Delignette-Muller, M. L. et al. fitdistrplus: An r package for fitting distributions. *J. Stat. Softw.* 64, 1–34 (2015).
72. Akaike, H. Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*, vol. 1, 199–213 (Springer, 1998).
73. Cullen, A. C., Frey, H. C. & Frey, C. H. Probabilistic techniques in exposure assessment: A handbook for dealing with variability and uncertainty in models and inputs (Springer Science & Business Media, 1999).