UNIVERSITÀ CATTOLICA DEL SACRO CUORE

INTERFACULTY ECONOMICS - BANKING, FINANCE AND INSURANCE SCIENCES
MASTER OF SCIENCE IN STATISTICAL AND ACTUARIAL SCIENCES
DATA ANALYTICS FOR BUSINESS AND ECONOMICS

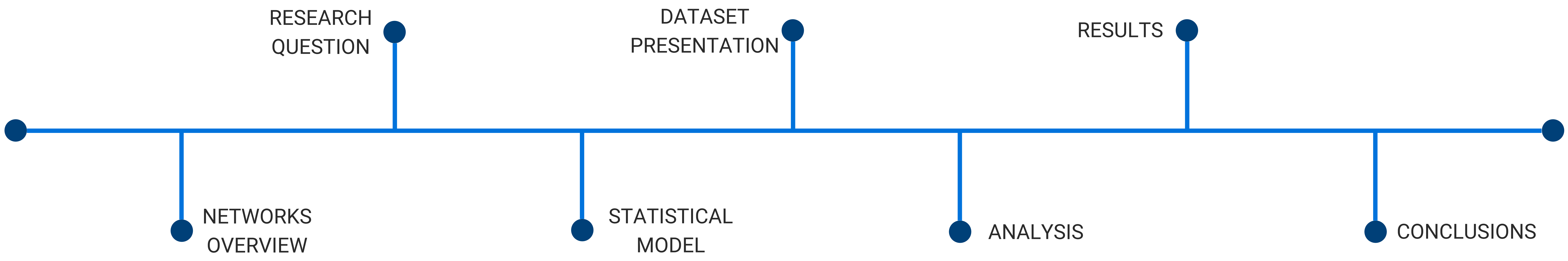
STATISTICAL MODELLING OF LONGITUDINAL NETWORK DATA: AN ANALYSIS OF THE TECH STOCK MARKET

Student: Andrea CORVI
ID: 4808460

Supervisor: Prof. Stefano PELUSO

Academic Year 2019-2020

Roadmap


RESEARCH QUESTION
DATASET PRESENTATION
RESULTS
NETWORKS OVERVIEW
STATISTICAL MODEL
ANALYSIS
CONCLUSIONS

Definition:

- A network ***N*** can be defined as a set of points (or nodes) and ties between them, $N = (V; E)$ (fig 1).

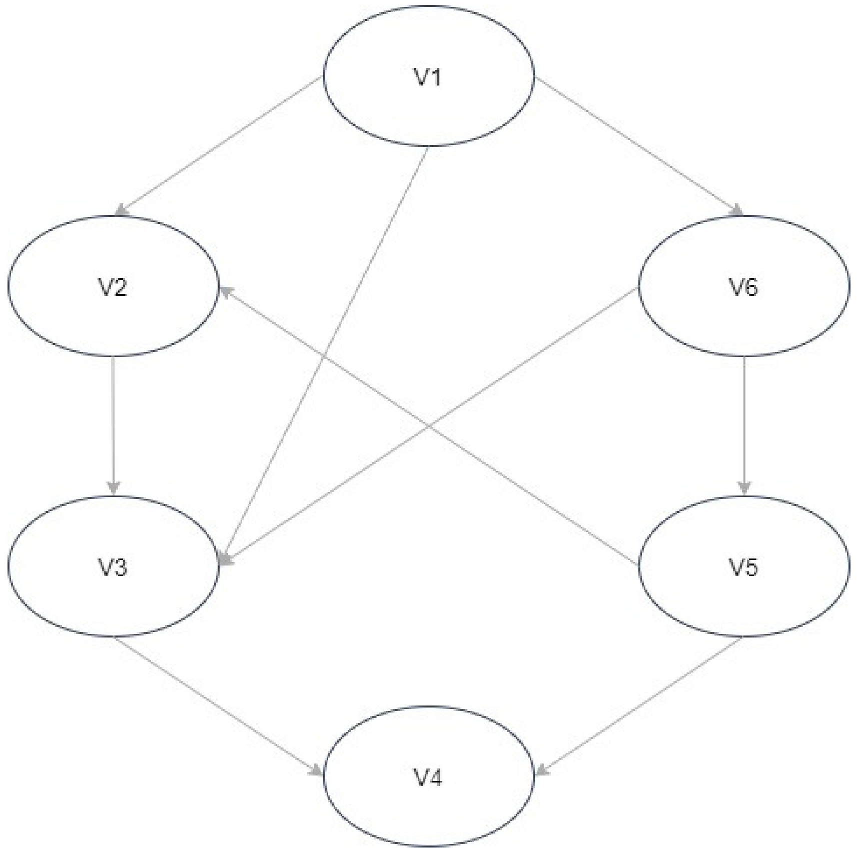


Fig.1 : Example of a network.

Possible representations:

- Adjacency *list* (fig. 2)
- Adjacency *matrix* (fig. 3).

Vertices	Ties
V1	V2, V3, V6
V2	V3
V3	V4
V4	None
V5	V2, V4
V6	V3, V5

Fig.2 : Example of adjacency list.

	V1	V2	V3	V4	V5	V6
V1	0	1	1	O	0	1
V2	0	0	1	0	0	0
V3	0	0	0	1	0	0
V4	0	0	0	0	0	0
V5	0	1	0	1	0	0
V6	0	0	1	0	1	0

Fig.3: Example of adjacency matrix.

Research Object

Longitudinal Networks

Research Question

To verify whether a statistical model developed for the analysis of social networks can be extended to a different field of application and specifically used to predict the monthly correlations between the returns of 13 tech companies at a specific time point.

Methodology

- Statistical model
- Data collection
- Dataset building

Statistical Model

The TERGM has the following **formulation**:

$$\mathcal{P}\left(N^t \mid N^{t-1}, \boldsymbol{\theta}\right) = \frac{1}{c\left(\boldsymbol{\theta}, N^{t-1}\right)} \exp\left\{\boldsymbol{\theta}' \boldsymbol{\Psi}\left(N^t, N^{t-1}\right)\right\}$$

Elements of the formulation:

- θ is the vector of the model coefficients.
- Ψ is a vector of statistics computed on the network *N*
- c is the normalizing constant.

Estimation:

1. Markov Chain Monte Carlo - MLE
2. Maximum Pseudolikelihood Estimation:

MPLE replaces likelihood with the product over conditional dyadic tie probabilities:

$$\pi_{ij}(\boldsymbol{\theta}) = \Pr\left(N_{ij} = 1 \mid \boldsymbol{N}_{-ij}, \boldsymbol{\theta}\right) = 1/\left[1 + \exp\left\{-\boldsymbol{\theta}'\delta_{ij}(\boldsymbol{\Psi}(\boldsymbol{N}))\right\}\right]$$

Advantages:

1. Faster than MCMC-MLE.
2. Converges asymptotically to the MLE as the number of network samples increases.

Data collection

- **13** major tech companies.
- **Daily stock returns** from 2005 to 2020.
- **Linear correlation** of the returns grouped by month.
- **Sharpe ratio** and **Beta** of the grouped returns.

↓

Correlations filtered according to **4 thresholds**:

- 0.6
- 0.65
- 0.7
- 0.75

If the correlation is greater than the threshold, then there is a tie between the two nodes.

Dataset building

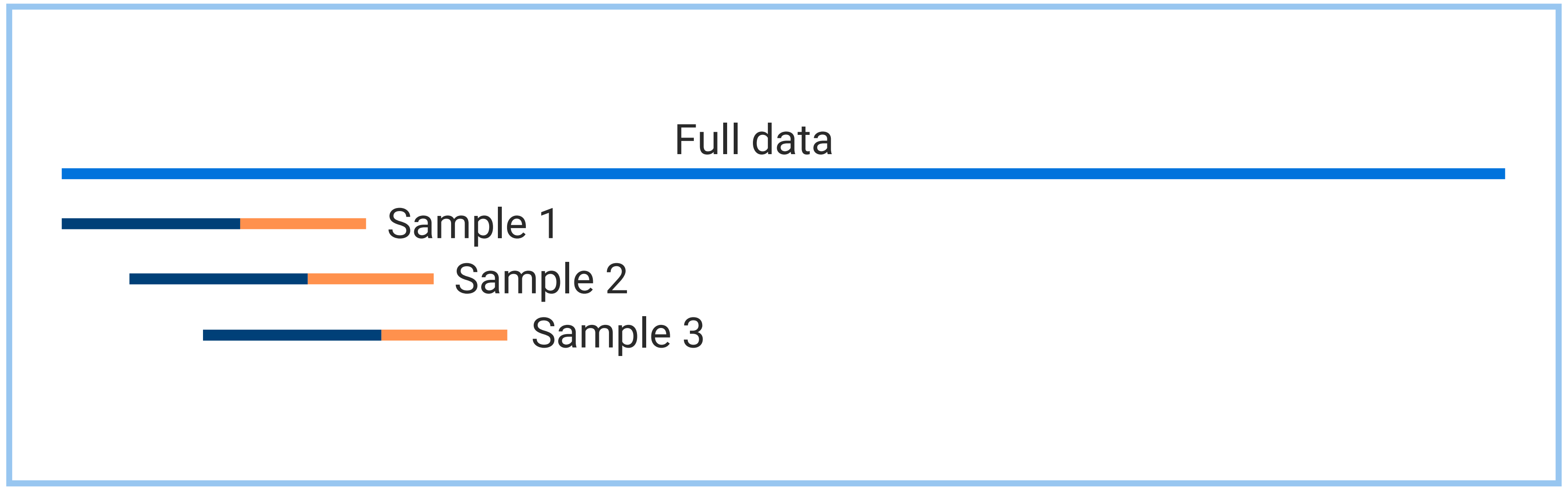
- **4 longitudinal networks** built according to the thresholds

↓ Each of them

- Represented through 13x13 adjacency matrices
- 188 time points.
- 13 nodes per network.
- Two nodal attributes: SR and β.

Rolling Origin Cross Validation

Model used to predict the status of the network at certain time point, given 3 years of data.



Results evaluated through average Roc AUC for each network

Out-of-sample Prediction

Performed on the most recent data, considering one year:

- *Training sample* = **11** time points (July 2019 - June 2020)
- *Test sample* = **1** time point (July 2020)

Predictive performance evaluated with Roc AUC.
Goodness of fit adressed with gof function:

1. Model fitted on training subset.
2. *n* simulations of the test subset are made.
3. Comparison of the simulations against the observed network.

Rolling Origin Cross Validation

The average Roc AUC for each network is:

- Network 1: 0.63
- Network 2: 0.63
- Network 3: 0.63
- Network 4: 0.65

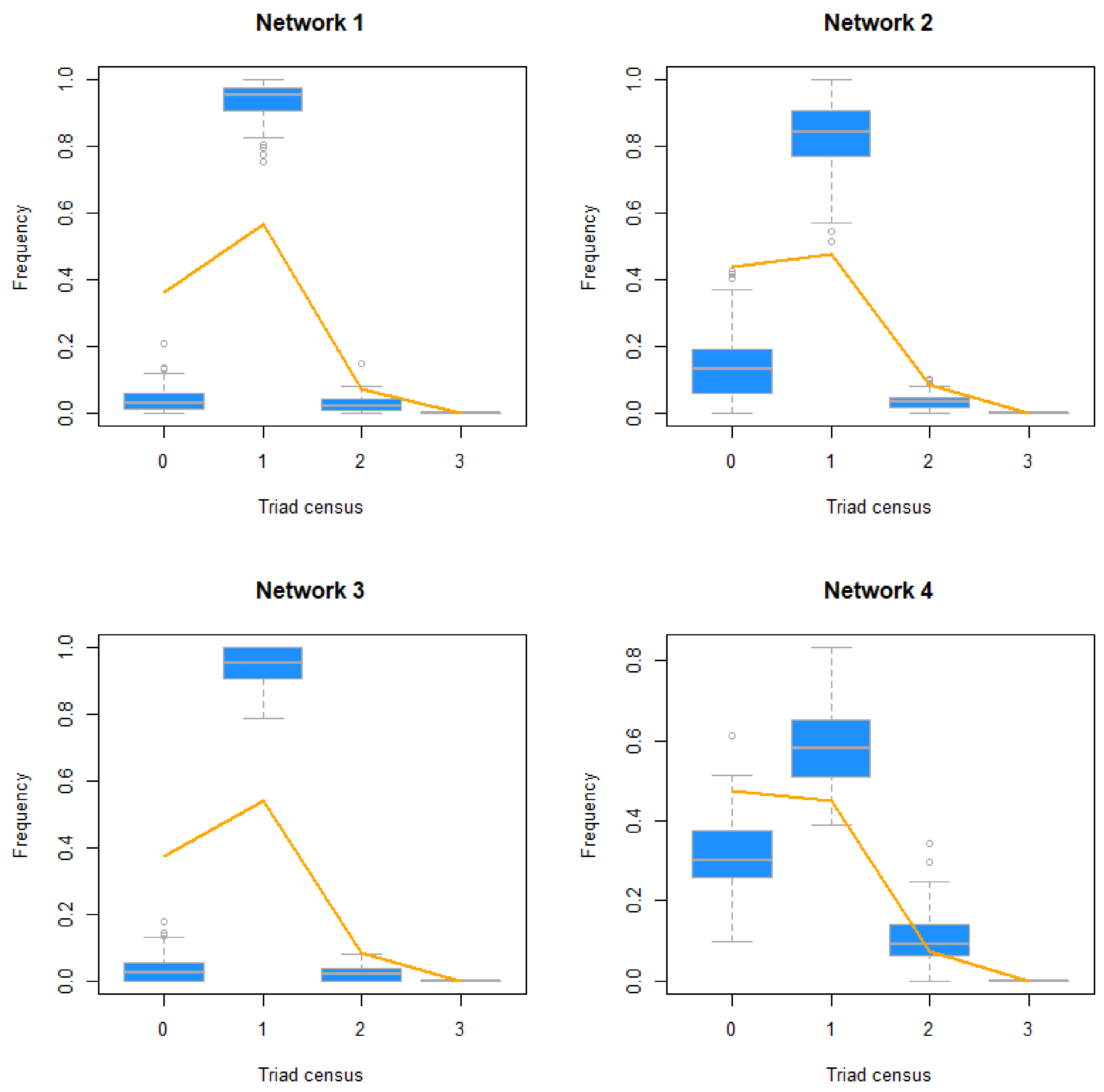
The model performs well in specific time windows but it is not very consistent, given the ROCV.

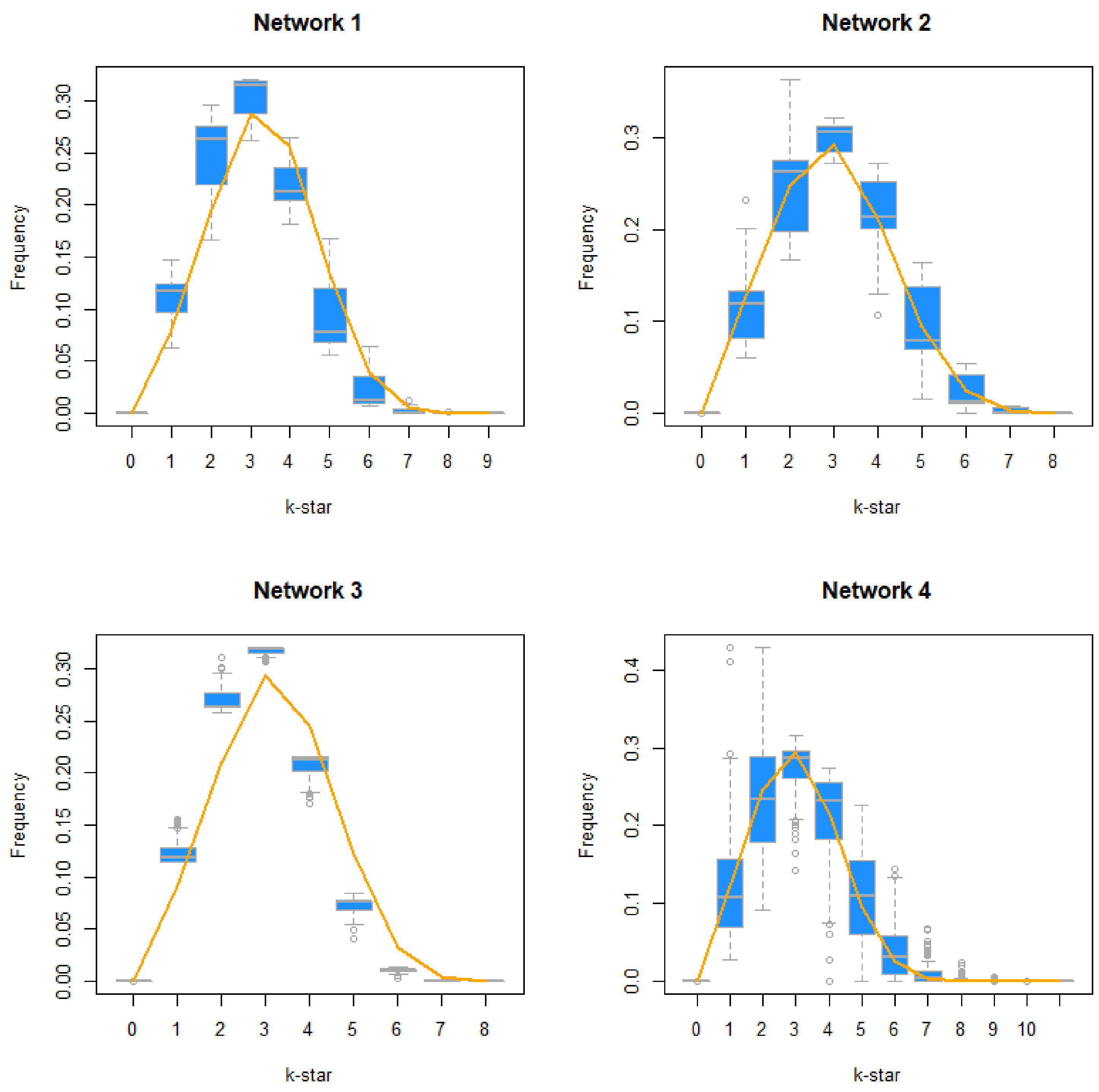
Out-of-sample Prediction

The Roc AUC for each network is:

- Network 1: 0.68
- Network 2: 0.75
- Network 3: 0.80
- Network 4: 0.78

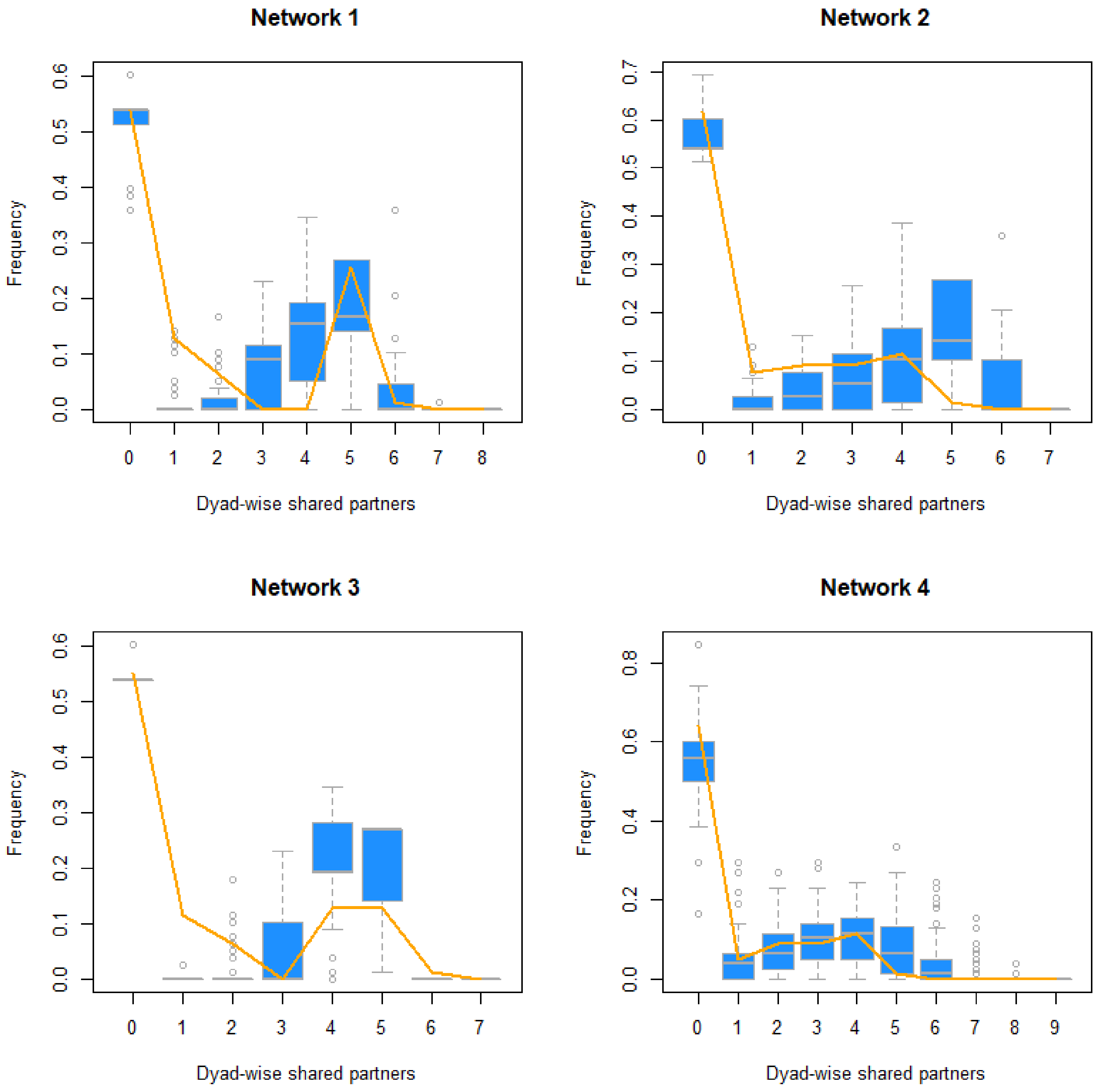
Goodness of fit of the "Out-of-sample Prediction" model.

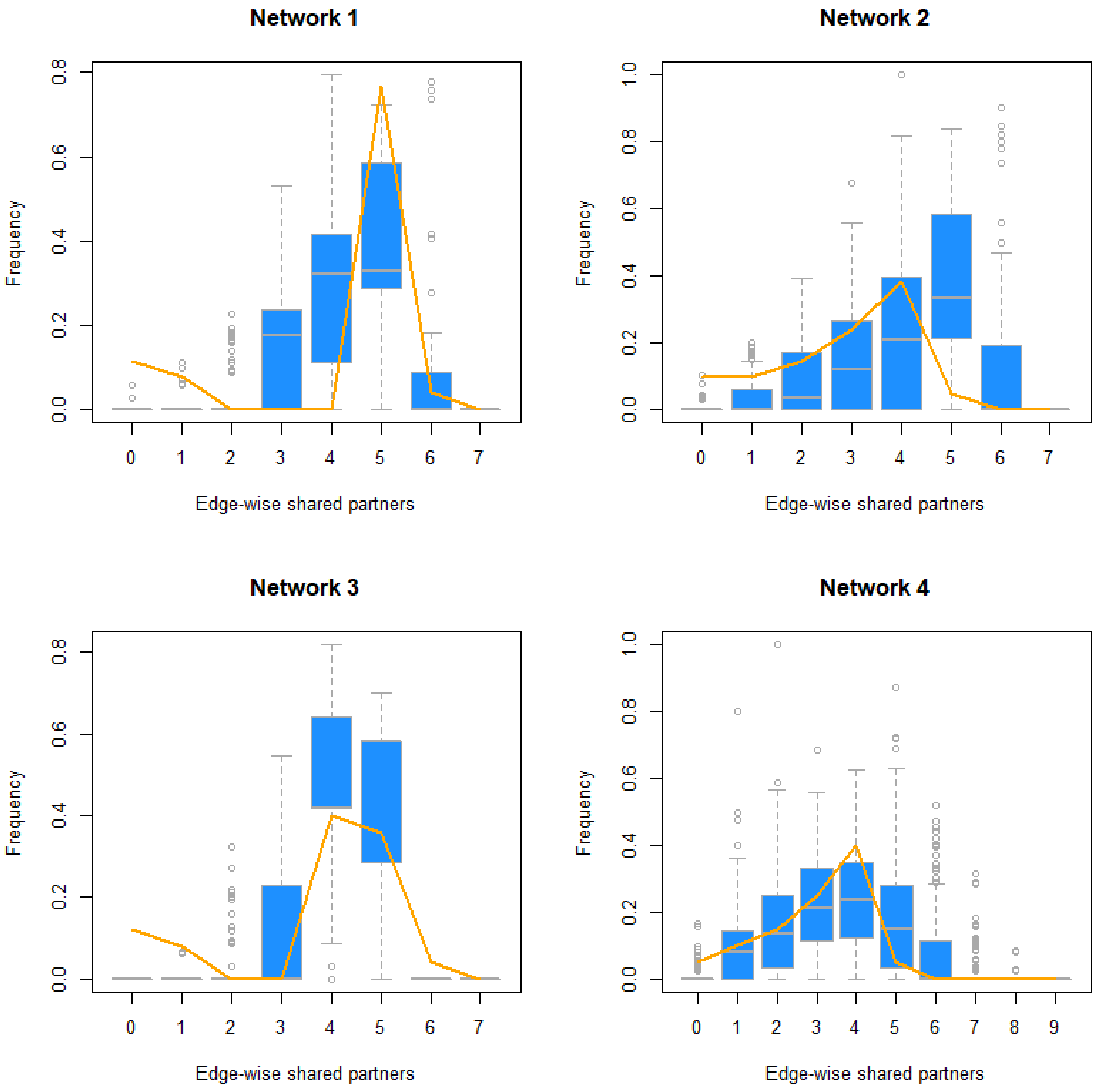




Andrea Corvi 4808460

Goodness of fit of the "Out-of-sample Prediction" model.





- Although the results of the **Rolling Origin** aren't consistent for each sample of the data, the performance on specific time frames is promising, as shown by the goodness of fit.

- Indeed, the results of the gof of the ***"Out-of-sample Prediction"*** model show that it is able to represent the original structure of the network, especially regarding specific statistics. Moreover it has obtained a **good overall predictive performance**.

Thus, the Temporal Exponential Random Graph model **can be extended** beyond its original field of application. However it is not consistent enough to perform well over long time span but it has to be fine tuned **to specific time frames**; in such applications the model can be exploited to predict the correlations between the nodes at specific time points.

- The financial world is by nature subject to changes depending both on endogenous and exogenous factors, which make it hard to have long-term predictions. The **networks** have been modelled using **mainly internal data**, the daily returns, and only partly using external data regarding the computing of the beta.
- This represents indeed a limit but also a **possible developement of the research**.

It is necessary to remark that the tech sector **evolves at a fast pace** which once more make it hard to have reliable long-term predictions.
In the following plots are represented two time steps of network 4: July 2019 and July 2020.
It is evident the change in the network structure:

- In 2019 there was a **triad** composed by the **japanese** companies which isn't present at one years distance.
- The position of Microsoft is peculiar since it isn't highly correlated with any company both in 2019 and in 2020.
- **Google is at the center** of a group in both the time steps.
- Also the position **outside of the main group** of **Amazon** in 2020 is rather interesting.

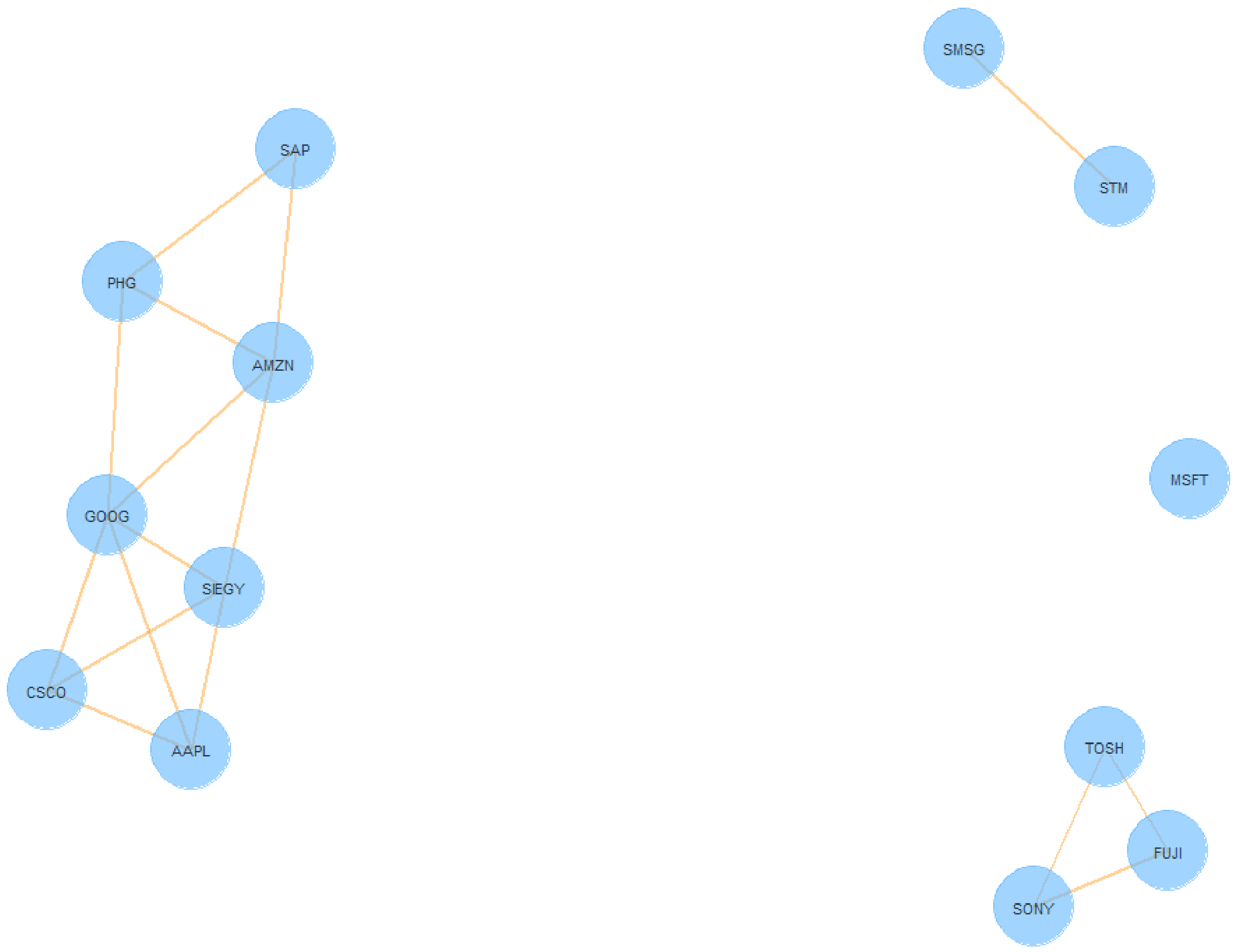


Fig.1: Network 4 at July 2019

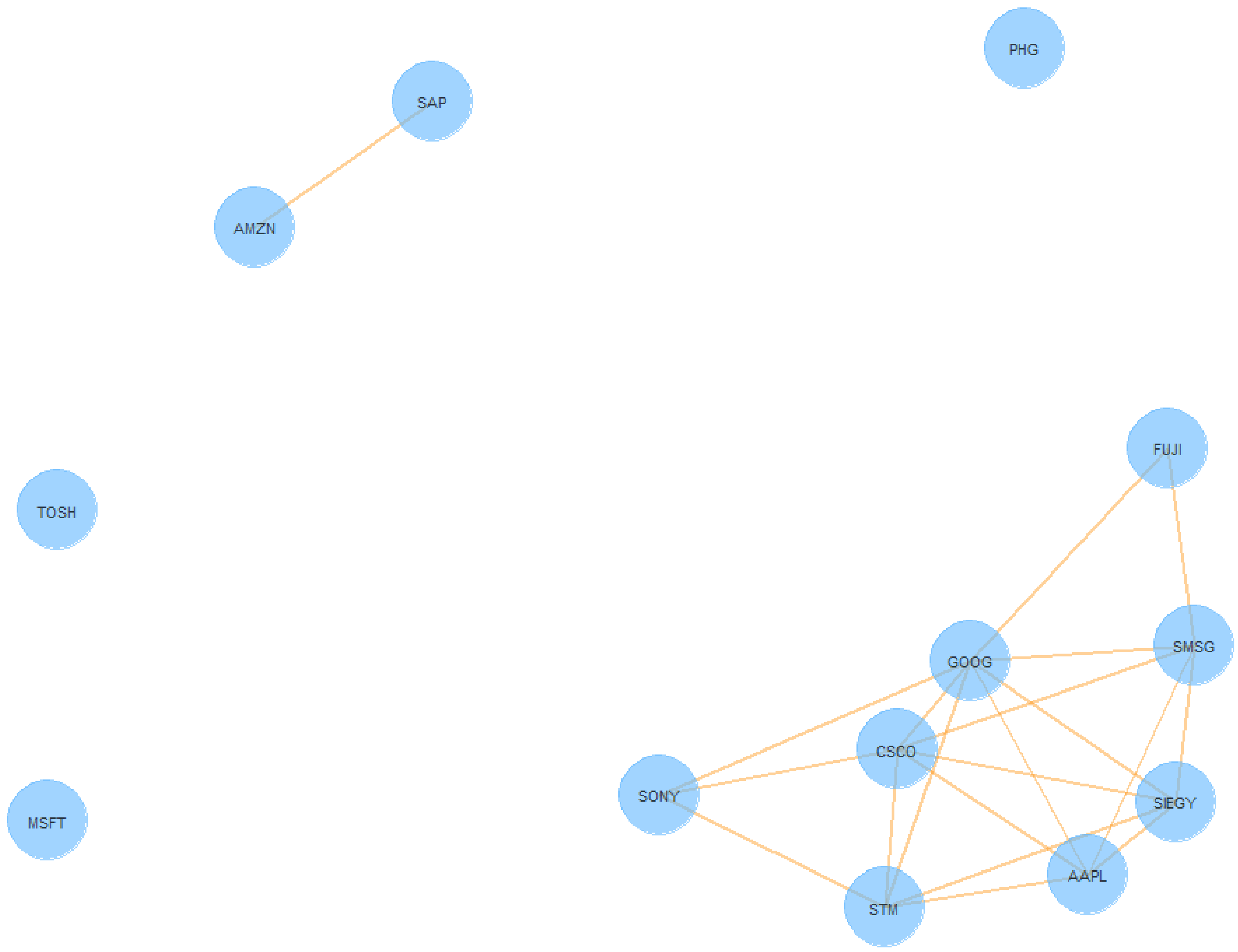


Fig.2: Network 4 at July 2020

Main Bibliography

- BIANCHI, FEDERICA, FRANCESCO BARTOLUCCI, STEFANO PELUSO, AND ANTONIETTA MIRA (2020). “LONGITUDINAL NETWORKS OF DYADIC RELATIONS USING LATENT TRAJECTORIES: EVIDENCE FROM THE EUROPEAN INTERBANK MARKET”. JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES C.
- DESMARAIS, B.A. AND S.J. CRANMER (2012). “STATISTICAL MECHANICS OF NETWORKS: ESTIMATION AND UNCERTAINTY”. PHYSICA A: STATISTICAL MECHANICS AND ITS APPLICATIONS 391.4, PP. 1865–1876.
- DURANTE, DANIELE AND DAVID DUNSON (OCT. 2014). “NONPARAMETRIC BAYES DYNAMIC MODELING OF RELATIONAL DATA”. BIOMETRIKA 101, PP. 883–898.
- ERDÖS, P. AND A. RÉNYI (1959). “ON RANDOM GRAPHS I”. PUBLICATIONES MATHEMATICAE DEBRECEN 6, P. 290.
- GOLDENBERG, ANNA, ALICE ZHENG, STEPHEN FIENBERG, AND EDOARDO AIROLDI (DEC. 2009). “A SURVEY OF STATISTICAL NETWORK MODELS”. FOUNDATIONS AND TRENDS IN MACHINE LEARNING 2..
- HANNEKE, STEVE, WENJIE FU, AND ERIC XING (2009). DISCRETE TEMPORAL MODELS OF SOCIAL NETWORKS.
- HOLLAND, PAUL W., KATHRYN BLACKMOND LASKEY, AND SAMUEL LEINHARDT (1983). “STOCHASTIC BLOCKMODELS: FIRST STEPS”. SOCIAL NETWORKS 5.2, PP. 109–137.
- KRIVITSKY, PAVEL AND MARK HANDCOCK (JAN. 2014). “A SEPARABLE MODEL FOR DYNAMIC NETWORKS”. JOURNAL OF THE ROYAL STATISTICAL SOCIETY. SERIES B, STATISTICAL METHODOLOGY 76, PP. 29–46.
- LEIFELD, PHILIP, SKYLER CRANMER, AND BRUCE DESMARAIS (2018). “TEMPORAL EXPONENTIAL RANDOM GRAPH MODELS WITH BTERGM: ESTIMATION AND BOOTSTRAP CONFIDENCE INTERVALS”. JOURNAL OF STATISTICAL SOFTWARE, ARTICLES 83.6, PP. 1–36.
- SNIJDERS,TOM(APR. 1996). “STOCHASTIC ACTOR-ORIENTED MODELS FOR NETWORK CHANGE”. JOURNAL OF MATHEMATICAL SOCIOLOGY - J MATH SOCIOL 21, PP. 149–172.
- STRAUSS, DAVID AND MICHAEL IKEDA (1990). “PSEUDOLIKELIHOOD ESTIMATION FOR SOCIAL NETWORKS”. JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION 85.409, PP. 204– 212.
- TASHMAN, LEONARD J. (2000). “OUT-OF-SAMPLE TESTS OF FORECASTING ACCURACY: AN ANALYSIS AND REVIEW”. INTERNATIONAL JOURNAL OF FORECASTING 16.4. THE M3- COMPETITION, PP. 437–450.
- WASSERMAN, STAN (MAR. 1987). “CONFORMITY OF TWO SOCIOMETRIC RELATIONS”. PSYCHOMETRIKA 52, PP. 3–18.
- WASSERMAN, STAN AND DAWN IACOBUCCI (FEB. 1988). “SEQUENTIAL SOCIAL NETWORK DATA”. PSYCHOMETRIKA 53, PP. 261–282.

THANK YOU FOR YOUR ATTENTION.